QUARTERLY OF APPLIED MATHEMATICS

Vol. V OCTOBER, 1947 No. 3

APPROXIMATIONS IN ELASTICITY BASED ON THE CONCEPT OF FUNCTION SPACE*

BY

W. PRAGER (*Brown University*) AND J. L. SYNGE (*Carnegie Institute of Technology*)

1. Introduction. The theory developed in the present paper applies to any elastic body (in general anisotropic) possessing a positive-definite strain-energy function, quadratic in the components of stress. This function provides us with a metric in a function space in which the point or vector represents a state of stress. The geometry of the function space follows Euclidean analogies closely, and is powerful in suggesting methods of approximation. The aim is to obtain approximate solutions of elastic boundary value problems with errors which are calculable, the error being measured in terms of distance in function space, or, equivalently, in terms of strain-energy.

After dealing with notation and basic concepts in Sec. 2, we discuss vectors in function space in Sec. 3. Sections 4 and 5 are intended to introduce the reader to the ideas which lie behind the general method; only a first approximation is discussed, and only the simplest types of boundary conditions.

In Sec. 6 more general types of boundary conditions are introduced, and higher approximations under these boundary conditions are treated in Secs. 7–11. Throughout, the basic plan is to locate the solution of the elastic boundary-value problem (considered as a point in function space) on a hypercircle of determinable center and radius. As a practical test of the method, it is used in Sec. 12 to obtain approximate solutions for the torsion of a prism of square cross section. In Sec. 13 reference is made to other work, and some known results are strengthened by use of the present method.

2. Notation and basic concepts. Latin suffixes take the range of values 1, 2, 3 and the summation convention operates on repeated suffixes. The coordinates x_i are rectangular cartesians, and differentiation with respect to a coordinate is indicated by a comma ($F_{,i}=\partial F/\partial x_i$).

By a *state* of an elastic body we understand a set of six stress components E_{ij} ($E_{ij}=E_{ji}$), given as functions of the coordinates throughout the body. For simplicity, we shall consider only functions which are continuous in the body and on its surface, and possess continuous first and second order derivatives in the body. It will be obvious that these conditions may be weakened, in the sense that these conditions hold throughout each of a finite number of parts into which the body is divided, with suitable continuity conditions across the surfaces which separate the parts.

The body is assumed to possess a *strain-energy function*

$$W = \tfrac{1}{2}c_{ijmn}E_{ij}E_{mn}, \tag{2.1}$$

* Received September 20, 1946.

where the constants c_{ijmn} satisfy

$$c_{ijmn} = c_{jimn} = c_{ijnm} = c_{mnij}, \tag{2.2}$$

and are such that the form W is positive-definite (i.e. it is positive unless all the stress components vanish).

We now introduce the *strain components* e_{ij} in an unconventional way. Instead of defining strain in terms of displacement, we define it by the generalized Hooke's law

$$e_{ij} = c_{ijmn}E_{mn}. \tag{2.3}$$

Thus (2.1) may be written

$$W = \tfrac{1}{2}e_{ij}E_{ij}. \tag{2.4}$$

It follows from (2.2) that for any two states, E_{ij} and E'_{ij}, we have the *reciprocity relation*

$$e_{ij}E'_{ij} = e'_{ij}E_{ij}. \tag{2.5}$$

The usual *equations of compatibility* read

$$e_{ij,mn} + e_{mn,ij} = e_{im,jn} + e_{jn,im}. \tag{2.6}$$

By (2.3), these can be translated into conditions on the stress components. In general, they will *not* be satisfied by an arbitrary state. But if they are satisfied, then the partial differential equations

$$u_{i,j} + u_{j,i} = 2e_{ij}, \tag{2.7}$$

have solutions u_i, unique to within an infinitesimal rigid body displacement,[1] and u_i is the *displacement* for the state E_{ij}. We are to remember that an arbitrary state has in general no corresponding displacement.

Should we wish to construct a state satisfying the equations of compatibility, the simplest plan is to choose a set of displacements, use (2.7) to obtain e_{ij}, and then solve (2.3) to obtain E_{ij}. The compatibility equations (2.6) will be automatically satisfied on account of (2.7).

Throughout this paper, we shall suppose that body forces are absent.* Thus the *equations of equilibrium* read

$$E_{ij,j} = 0. \tag{2.8}$$

In general these equations will not be satisfied by an arbitrary state.

We shall denote by n_i the unit vector normal to the surface of the body, pointing outward. Then the stress across the surface is

$$T_i = E_{ij}n_j. \tag{2.9}$$

The following notation for "inner products" will be found convenient. Since it will be obvious whether an integration throughout the body or an integration over its surface is implied, one type of notation serves for both types of inner product. We shall write

---

[1] I. S. Sokolnikoff, Mathematical Theory of Elasticity, McGraw-Hill, New York, 1946, p. 24.

* The extension to the case where body forces are present will be made in a later paper.

$$(e \cdot E') = \int e_{ij} E'_{ij} dv,$$
$$(u \cdot T') = \int u_i T'_i dS, \tag{2.10}$$

the former integral being taken throughout the body, and the latter over its surface. The interpretation of similar expressions will be obvious to the reader.

In attempting to solve a problem in elasticity, we seek a state which satisfies

(*i*) the equations of compatibility (2.6),

(*ii*) the equations of equilibrium (2.8),

(*iii*) the assigned boundary conditions, which will generally be T_i assigned, or u_i assigned, or T_i assigned in part and u_i assigned in part.

We approach the solution by considering states in which one or more of these conditions are relaxed.

3. Geometry in the function spaces of states. Since the power of the present method lies largely in the stimulation of geometrical intuition, it is well to pictorialize from the start. We shall develop some simple properties of the function space of states.

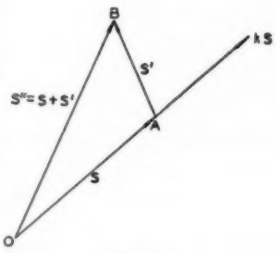

FIG. 1.

The unstressed state, $E_{ij} = 0$, is represented by the *origin* O (Fig. 1). Any other state is represented by a *point* such as A. Just as we describe the position of a point in ordinary space by a position-vector drawn from the origin, so we describe the point or state A by the *vector* $\overrightarrow{OA}$, or more compactly by a single letter (**S**) in heavy type. Thus the symbol **S** represents a state of stress (six functions of position throughout the body).

We *add* vectors in an obvious way. If we have two states,

S with stress components E_{ij}, and

S' with stress components E'_{ij},

we define the sum $\mathbf{S''} = \mathbf{S} + \mathbf{S'}$ to be the state with stress components

$$E''_{ij} = E_{ij} + E'_{ij}.$$

The definition of $\mathbf{S} - \mathbf{S'}$ is obvious.

To multiply a vector by a scalar constant k, we write $\mathbf{S'} = k\mathbf{S}$, and define **S'** by the equations $E'_{ij} = kE_{ij}$. Note that E_{ij} are functions of position in the body, but k is a constant. The distributive laws are satisfied, so that

$$(k + k')\mathbf{S} = k\mathbf{S} + k'\mathbf{S}, \qquad k(\mathbf{S} + \mathbf{S'}) = k\mathbf{S} + k\mathbf{S'}.$$

So far we have mentioned only vectors drawn out from the origin O, i.e. *bound vectors*. But the idea of a *free vector*, familiar in ordinary space, can be used with advantage. Apart from the geometrical representation, there is no real distinction between a free vector and a vector bound to O: each corresponds to a given set of stress components. But in the geometrical representation, a free vector **S** starts from *any* initial state A and proceeds to that state B for which the stress components exceed those of A by the stress components corresponding to **S**. In fact, a free vector may be

regarded as a transition from one state to another, and we make it a bound vector if we specify the initial state.

The algebraic procedures apply to free vectors. The addition of vectors follows the familiar parallelogram law (Fig. 1).

It should be amply clear from the context or from the appropriate diagram whether a free or bound vector is meant. In any case of doubt, a vector should be interpreted as a vector drawn from the origin O.

Fig. 1 will also draw the reader's attention to the geometrical relation between the vectors $\mathbf{S}$ and $k\mathbf{S}$. If k were negative, there would be a reversal of sense.

We define the *length* or *magnitude* S of a vector $\mathbf{S}$ by

$$S^2 = \int 2W dv = \int e_{ij}E_{ij}dv = (e\cdot E), \qquad S \geqq 0, \tag{3.1}$$

the integrals being taken throughout the body. Here e_{ij} is the strain corresponding to the stress E_{ij}, by (2.3). Since the strain-energy W is positive-definite, the length of any vector is real, and is zero only for the unstressed state, $E_{ij}=0$. Sometimes it will be convenient to denote the length of a vector $\mathbf{S}$ by $|\mathbf{S}|$. A vector $\mathbf{S}$ is a *unit vector* if $S=1$.

The distance between the extremities of two vectors, $\mathbf{S}$ and $\mathbf{S}'$, drawn from O, is defined to be the length of the vector $\mathbf{S}-\mathbf{S}'$. Thus this distance is given by

$$|\mathbf{S}-\mathbf{S}'|^2 = \int (e_{ij}-e'_{ij})(E_{ij}-E'_{ij})dv. \tag{3.2}$$

It vanishes if, and only if, $\mathbf{S}=\mathbf{S}'$, i.e. if the two states are identical.

The *scalar product* plays an important role. We define the scalar product of two vectors as the inner product of the corresponding states:

$$\mathbf{S}\cdot\mathbf{S}' = \int e_{ij}E'_{ij}dv = (e\cdot E'). \tag{3.3}$$

It is obvious, from the reciprocity relation (2.5), that the scalar product has the commutative property

$$\mathbf{S}\cdot\mathbf{S}' = \mathbf{S}'\cdot\mathbf{S}. \tag{3.4}$$

Two vectors are said to be *perpendicular* or *orthogonal* if their scalar product vanishes.

Let us explore the physical meaning of the scalar product. Let us suppose that $\mathbf{S}$ satisfies the equations of compatibility, so that it arises from a displacement, and that $\mathbf{S}'$ satisfies the equations of equilibrium. Then

$$\begin{aligned}\mathbf{S}\cdot\mathbf{S}' = (e\cdot E') &= \int e_{ij}E'_{ij}dv = \int u_{i,j}E'_{ij}dv \\ &= \int u_iE'_{ij}n_j dS - \int u_iE'_{ij,j}dv,\end{aligned} \tag{3.5}$$

by Green's theorem. The last integral vanishes, since $\mathbf{S}'$ satisfies the equations of equilibrium (2.8). Hence

$$\mathbf{S}\cdot\mathbf{S}' = \int u_iT'_i\,dS = (u\cdot T'); \tag{3.6}$$

in words, *the scalar product* $\mathbf{S}\cdot\mathbf{S}'$ *equals the work done in the displacement of* $\mathbf{S}$ *by the surface stress of* $\mathbf{S}'$. If the vectors are orthogonal, this work is zero.

As an example, consider a prismatic bar. Let the states $\mathbf{S}$ and $\mathbf{S}'$ correspond to simple tension and flexure, respectively. Then $\mathbf{S}$ is orthogonal to $\mathbf{S}'$.

It should be noted that the scalar product can be interpreted in terms of work only if one of the states satisfies the equations of equilibrium and the other the equations of compatibility.

We come now to a fundamental inequality. From the positive-definite character of strain-energy, it follows that if $\mathbf{S}$ and $\mathbf{S}'$ are any two arbitrary states, and k an arbitrary real number, then

$$\int (e'_{ij} - ke_{ij})(E'_{ij} - kE_{ij})dv \geqq 0, \tag{3.7}$$

the integral being taken throughout the body. This reduces to

$$S'^2 - 2k\mathbf{S}\cdot\mathbf{S}' + k^2S^2 \geqq 0, \tag{3.8}$$

and since this holds for all real values of k, we deduce the Schwarzian inequality

$$|\mathbf{S}\cdot\mathbf{S}'| \leqq SS'. \tag{3.9}$$

If the sign of equality holds in (3.9), then it is easy to see that the states $\mathbf{S}$ and $\mathbf{S}'$ must be connected by a relation $\mathbf{S}'=K\mathbf{S}$, where K is some real constant; in other words, the vectors $\mathbf{S}$ and $\mathbf{S}'$ have the same direction, or opposite directions.

We can now define the *angle* θ between two vectors $\mathbf{S}$ and $\mathbf{S}'$ by

$$\cos\theta = \frac{\mathbf{S}\cdot\mathbf{S}'}{SS'}, \quad 0 \leqq \theta \leqq \pi. \tag{3.10}$$

By virtue of (3.9), the angle so defined is always real, and is of course equal to $\frac{1}{2}\pi$ when the vectors are orthogonal. For the angle between two unit vectors, $\mathbf{I}$ and $\mathbf{I}'$, we have

$$\cos\theta = \mathbf{I}\cdot\mathbf{I}'. \tag{3.11}$$

The whole of Euclidean geometry holds in any linear subspace of our function space based on a finite number of vectors. In particular, we have the theorem of Pythagoras, the theorem that any side of a triangle is less than the sum of the other two sides, and the theorem that the greater side of a triangle is opposite the greater angle. All this is well known, but the abstract and general form usually given to the theory of function space may easily obscure the intuitive simplicity of the approach.

Scalar multiplication is distributive, as is easily seen, so that

$$\mathbf{S}_1\cdot(\mathbf{S}_2 + \mathbf{S}_3) = \mathbf{S}_1\cdot\mathbf{S}_2 + \mathbf{S}_1\cdot\mathbf{S}_3. \tag{3.12}$$

As an exercise, let us prove the theorem of Pythagoras. We have, for any two vectors, $\mathbf{S}$ and $\mathbf{S}'$,

$$\begin{aligned} |\mathbf{S} - \mathbf{S}'|^2 &= (\mathbf{S} - \mathbf{S}')\cdot(\mathbf{S} - \mathbf{S}') \\ &= S^2 + S'^2 - 2\mathbf{S}\cdot\mathbf{S}' \\ &= S^2 + S'^2 - 2SS'\cos\theta, \end{aligned} \tag{3.13}$$

where θ is the angle between $\mathbf{S}$ and $\mathbf{S}'$. Putting $\theta = \frac{1}{2}\pi$, we get the theorem of Pythagoras.

4. First approximation with surface stress given. Let us suppose that the surface stress T_i is assigned, satisfying the conditions of statical equilibrium, so that a solution $\mathbf{S}$ exists. We may refer to the solution $\mathbf{S}$ as the *natural* state, to distinguish it from the artificial states which we shall introduce.

Let $\mathbf{S}^*$ be a state which satisfies the equations of equilibrium and the boundary conditions, but not the equations of compatibility. (If the equations of compatibility were satisfied, we would have $\mathbf{S}^* = \mathbf{S}$, the natural state.) Let us evaluate the scalar product $\mathbf{S} \cdot \mathbf{S}^*$. We find

$$\mathbf{S} \cdot \mathbf{S}^* = (e \cdot E^*) = \int e_{ij} E^*_{ij} dv = \int u_{i,j} E^*_{ij} dv$$

$$= \int u_i T^*_i dS = \int u_i T_i dS = (e \cdot E) = S^2; \tag{4.1}$$

in carrying this out, we have used the fact that $\mathbf{S}^*$ satisfies the equations of equilibrium, and also the fact that it satisfies the boundary conditions, so that $T_i^* = T_i$. Since $S^2 = \mathbf{S} \cdot \mathbf{S}$, we may write (4.1) in the form

$$\mathbf{S} \cdot (\mathbf{S} - \mathbf{S}^*) = 0. \tag{4.2}$$

This tells us that the vectors $\mathbf{S}$ and $\mathbf{S} - \mathbf{S}^*$ are orthogonal; *this means that the extremity of* $\mathbf{S}$ *is located on a hypersphere* H *having the vector* $\mathbf{S}^*$ *for diameter*; in fact, (4.2) is the equation of this hypersphere, $\mathbf{S}$ being regarded as a current vector. The center of the hypersphere is at $\frac{1}{2}\mathbf{S}^*$, and its radius is $\frac{1}{2}S^*$ (Fig. 2). The equation (4.2) may also be written in the equivalent form

$$\left| \mathbf{S} - \tfrac{1}{2}\mathbf{S}^* \right| = \tfrac{1}{2}S^*, \tag{4.3}$$

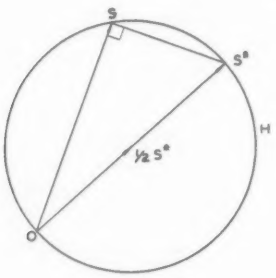

FIG. 2.

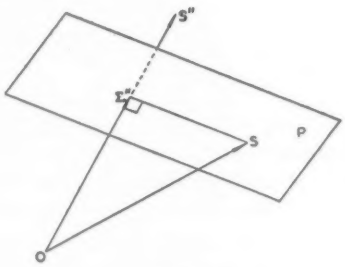

FIG. 3.

which shows up the center and the radius.

We now take another state $\mathbf{S}''$ which satisfies the equations of compatibility, but not in general either the equations of equilibrium or the boundary conditions. In fact, we can build such a vector $\mathbf{S}''$ by choosing an arbitrary set of displacements throughout the body. Let us investigate the scalar product $\mathbf{S} \cdot \mathbf{S}''$, $\mathbf{S}$ being the natural state as before. It is easy to see that

$$\mathbf{S} \cdot \mathbf{S}'' = (e'' \cdot E) = (u'' \cdot T). \tag{4.4}$$

Although $\mathbf{S}$ is unknown, we do know T_i from the assigned boundary conditions, and so the expression $(u'' \cdot T)$ is calculable. If θ is the angle between $\mathbf{S}$ and $\mathbf{S}''$, we have

$$S \cos\theta = \mathbf{S} \cdot \mathbf{S}''/S'' = (u'' \cdot T)/S''. \tag{4.5}$$

The right hand side is a definite calculable number, and $S \cos\theta$ is the orthogonal projection of $\mathbf{S}$ on $\mathbf{S}''$. It follows that the extremity of the natural vector $\mathbf{S}$ lies on a definite hyperplane P which is orthogonal to $\mathbf{S}''$ (Fig. 3).

Now

$$\mathbf{S}^* \cdot \mathbf{S}'' = (e'' \cdot E^*) = (u'' \cdot T^*) = (u'' \cdot T). \tag{4.6}$$

Comparison with (4.4) reveals a rather remarkable fact: *the extremity of the vector* $\mathbf{S}^*$ *lies in the hyperplane* P.

On account of (4.6), we can write (4.4) in the form

$$\mathbf{S} \cdot \mathbf{S}'' = \mathbf{S}^* \cdot \mathbf{S}''; \tag{4.7}$$

thinking of $\mathbf{S}$ as a current vector, we may regard this as the equation of the hyperplane P.

We have now located the extremity of the natural vector $\mathbf{S}$ on the hypersphere (4.2) and the hyperplane (4.7); we have therefore located it on the *hypercircle* Γ which is the intersection of the hypersphere and the hyperplane. The situation is therefore as shown schematically in Fig. 4. The hypercircle Γ passes through the extremity of $\mathbf{S}^*$ and also through Σ'', the foot of the perpendicular dropped from O on P. We have

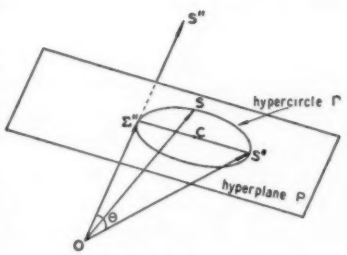

FIG. 4.

$$\mathbf{\Sigma}'' = \mathbf{S}'' \frac{\mathbf{S}^* \cdot \mathbf{S}''}{S''^2}. \tag{4.8}$$

If $\mathbf{I}''$ is a unit vector codirectional with $\mathbf{S}''$ (and therefore also satisfies the equations of compatibility), we have

$$\mathbf{I}'' = \mathbf{S}''/S'',$$

and (4.8) can be written in the form

$$\mathbf{\Sigma}'' = \mathbf{I}''(\mathbf{S}^* \cdot \mathbf{I}''). \tag{4.9}$$

The points $\mathbf{\Sigma}''$ and $\mathbf{S}^*$ are at the ends of a diameter of the hypercircle Γ, because

$$(\mathbf{S} - \mathbf{\Sigma}'') \cdot (\mathbf{S} - \mathbf{S}^*) = 0$$

as follows from (4.2) and (4.7). The center $\mathbf{C}$ of Γ is therefore at

$$\mathbf{C} = \tfrac{1}{2}(\mathbf{S}^* + \mathbf{\Sigma}'') = \tfrac{1}{2}[\mathbf{S}^* + \mathbf{I}''(\mathbf{S}^* \cdot \mathbf{I}'')], \tag{4.10}$$

and the radius R of Γ is given by

$$R = \tfrac{1}{2}(S^{*2} - \Sigma''^2)^{1/2} = \tfrac{1}{2}[S^{*2} - (\mathbf{S}^* \cdot \mathbf{I}'')^2]^{1/2}. \tag{4.11}$$

Of all points on Γ, the endpoint of $\mathbf{\Sigma}''$ is nearest to O and the endpoint of $\mathbf{S}^*$ most distant from O. Thus, $\Sigma''^2 \leqq S^2 \leqq S^{*2}$ or

$$(\mathbf{S}^* \cdot \mathbf{I}'')^2 \leqq S^2 \leqq S^{*2}. \tag{4.12}$$

This relation gives upper and lower bounds for the strain energy associated with the state $\mathbf{S}$.

We may also write (4.11) in the form

$$R = \tfrac{1}{2} S^* \sin \theta, \tag{4.13}$$

where θ is the angle between $\mathbf{S}^*$ and $\mathbf{S}''$.

Let us now consider the question of getting an approximation to the solution $\mathbf{S}$ of the given boundary value problem. Various definitions of a "good" approximation might be given. We might base a definition on the maximum deviation of displacement or stress from the correct value. We shall, however, use our geometrical picture as the basis of a test for goodness, and say that *an approximate solution* $\bar{\mathbf{S}}$ *is good if the distance (in function space) between* $\bar{\mathbf{S}}$ *and the true or natural solution* $\mathbf{S}$ *is small.* We shall define the *error* ϵ of any approximate solution by

$$\epsilon = |\bar{\mathbf{S}} - \mathbf{S}|.$$

Returning to the discussion of the hypercircle Γ, we ask whether our knowledge based on $\mathbf{S}^*$ and $\mathbf{S}''$ enables us to give an approximate solution. Suppose we accepted $\mathbf{S}^*$ as an approximate solution, how great an error would we commit? It is easy to set an upper bound to this error, because $\mathbf{S}$ lies on Γ, and the point on Γ most distant from $\mathbf{S}^*$ is the point $\mathbf{\Sigma}''$, which is diametrically opposed and at a distance $2R$. Thus we may state that *the error of* $\mathbf{S}^*$ *does not exceed* $2R$, where R is given by (4.11).

Similarly, if $\mathbf{\Sigma}''$, as in (4.8) or (4.9), is taken as an approximate solution, the error does not exceed $2R$.

We can however do better. *If we take* $\mathbf{C}$, *the center of the hypercircle* Γ, *as approximate solution, the error is precisely* R. Thus, if we change from $\mathbf{S}^*$ to $\mathbf{C}$, we exchange an unknown error which *may* be as great as $2R$ for a *certain* error which is only R.

It should be pointed out that the approximate solution $\mathbf{C}$ will, in general, not satisfy the equations of compatibility, nor the equations of equilibrium, nor the boundary conditions.

The hypercircle Γ, formed for states $\mathbf{S}^*$ and $\mathbf{S}''$, gives us a yardstick by which we can assess any proposed approximate solution $\bar{\mathbf{S}}$. If D_1 is the least distance of $\bar{\mathbf{S}}$ from Γ and D_2 the greatest distance, then the error ϵ of $\bar{\mathbf{S}}$ satisfies

$$D_1 \leqq \epsilon \leqq D_2. \tag{4.14}$$

It is easy to prove by elementary trigonometry that D_1^2 and D_2^2 are given by

$$(\bar{\mathbf{S}} - \mathbf{C})^2 + R^2 \mp 2R\,[(\bar{\mathbf{S}} - \mathbf{C})^2 - (\mathbf{I}'' \cdot (\bar{\mathbf{S}} - \mathbf{C}))^2]^{1/2}. \tag{4.15}$$

5. First approximation with surface displacement given. In the preceding section we have considered the case where the surface stress T_i is given. We shall now take up the other type of problem in which the displacement u_i is given on the bounding surface of the body. The argument runs closely parallel to that of Sec. 4.

First we choose a state $\mathbf{S}^*$ which satisfies the equations of compatibility and the boundary conditions. This is a much easier task than we had in Sec. 4, because

all we have to do is to take a system of displacements which assume assigned values on the bounding surface, and satisfy the general conditions regarding continuity of derivatives.

Next we choose a state $\mathbf{S}''$ which satisfies the equations of equilibrium, but not in general the equations of compatibility, so that satisfaction of the boundary conditions is meaningless, since displacement does not exist. Let $\mathbf{I}''$ be the corresponding unit vector.

We have

$$\mathbf{S}\cdot\mathbf{S}^* = (e^*\cdot E) = (u^*\cdot T) = (u\cdot T) = S^2. \tag{5.1}$$

Accordingly, the endpoint of $\mathbf{S}$ lies on the hypersphere

$$\mathbf{S}\cdot(\mathbf{S} - \mathbf{S}^*) = 0. \tag{5.2}$$

Also,

$$\mathbf{S}\cdot\mathbf{S}'' = (e\cdot E'') = (u\cdot T'') \tag{5.3}$$

which is calculable, and

$$\mathbf{S}^*\cdot\mathbf{S}'' = (e^*\cdot E'') = (u^*\cdot T'') = (u\cdot T''). \tag{5.4}$$

Thus,

$$\mathbf{S}\cdot\mathbf{S}'' = \mathbf{S}^*\cdot\mathbf{S}''. \tag{5.5}$$

Let us compare the equations (5.2) and (5.5) with (4.2) and (4.7), respectively. They are formally the same equations, although $\mathbf{S}^*$ and $\mathbf{S}''$ now have new meanings. Hence, just as we obtained (4.10), (4.11) and (4.12), we are now led to the result: *when the surface displacement is given, the natural vector* $\mathbf{S}$ *has its endpoint on a hypercircle with the center at*

$$\mathbf{C} = \tfrac{1}{2}[\mathbf{S}^* + \mathbf{I}''(\mathbf{S}^*\cdot\mathbf{I}'')], \tag{5.6}$$

and the radius R *given by*

$$R = \tfrac{1}{2}[S^{*2} - (\mathbf{S}^*\cdot\mathbf{I}'')^2]^{1/2}; \tag{5.7}$$

the strain energy associated with the state $\mathbf{S}$ *is bounded in accordance with*

$$(\mathbf{S}^*\cdot\mathbf{I}'')^2 \leqq S^2 \leqq S^{*2}. \tag{5.8}$$

6. Boundary conditions. Two types of boundary conditions may be regarded as fundamental: (i) stress assigned, (ii) displacement assigned. As we have seen in Secs. 4 and 5, there is remarkable duality between these two types of boundary condition. This duality extends to more general types of boundary condition, which will now be described before we proceed to generalize the ideas developed in the preceding sections.

We consider boundary conditions which involve the following elements at points of the bounding surface:

normal stress $T_{(n)}$; tangential stress $T_{(t)}$;

normal displacement $u_{(n)}$; tangential displacement $u_{(t)}$.

The normal elements are of course scalars and the tangential elements vectors in the tangent plane.

We shall classify points on the bounding surface according to the conditions assigned at them. Thus, a point will be said to be of the class $[u_{(n)}, T_{(t)}]$ if the normal displacement and the tangential stress are assigned at the point. Cases of vanishing elements are important, and for them we shall use the following type of notation: $[u_{(n)}=0, T_{(t)}]$ means that the normal displacement and the tangential stress are assigned, and that the normal displacement is zero.

Two types of boundary condition will now be defined. These may be regarded as generalizations of the simple types of boundary conditions in which stress is assigned all over the surface, or displacement is assigned all over the surface.

STRESS BOUNDARY CONDITIONS (or briefly SBC): The bounding surface may be divided into regions such that all the points in any region belong to one of the following classes:

$$[T_{(n)}, T_{(t)}], \quad [T_{(n)}, u_{(t)}=0], \quad [T_{(t)}, u_{(n)}=0], \quad [u_{(n)}=0, u_{(t)}=0]. \tag{6.1}$$

Obviously, if the complete surface stress T_i is assigned all over the surface, all points belong to the first class, and so this is a particular case of SBC. Note that in SBC any displacement elements which may be assigned are zero.

DISPLACEMENT BOUNDARY CONDITIONS (or briefly DBC): The bounding surface may be divided into regions such that all the points in any region belong to one of the following classes:

$$[u_{(n)}, u_{(t)}], \quad [u_{(n)}, T_{(t)}=0], \quad [u_{(t)}, T_{(n)}=0], \quad [T_{(n)}=0, T_{(t)}=0]. \tag{6.2}$$

Note that the case where displacement is assigned all over the surface is a particular case of DBC, and that in DBC any stress elements which may be assigned are zero.

As an illustration, consider a horizontal beam carrying a load on its smooth upper surface and supported on smooth unyielding supports. The points of the surface may be classified as follows:

upper surface under load $[T_{(n)}, T_{(t)}=0]$
surface in contact with supports $[T_{(t)}=0, u_{(n)}=0]$
rest of surface $[T_{(n)}=0, T_{(t)}=0]$

The boundary conditions are evidently SBC, only the first and third classes of (6.1) occurring.

As a second illustration, consider the classical Saint Venant torsion problem for a beam of arbitrary section. The classification is as follows:

ends of beam $[u_{(t)}, T_{(n)}=0]$
sides of beam $[T_{(n)}=0, T_{(t)}=0]$

Here we have DBC, only the third and fourth classes of (6.2) occurring.

7. Associated and complementary differential equations and states. When we have to deal with SBC, we have reason to regard the equations of equilibrium, rather than the equations of compatibility, as particularly associated with the boundary conditions. On the other hand, when we have to deal with DBC, we regard the equations of compatibility as particularly associated with the boundary conditions. The real reason for these associations lies in the theory which follows, but we may note the following facts which will help in remembering which differential equations are associated with which boundary conditions.

In the simplest type of SBC ($E_{ij}n_j$ assigned) and in the equations of equilibrium ($E_{ij,j}=0$), the stress components are involved in a very simple way. On the other

hand, the assignment of DBC would be meaningless unless the equations of compatibility were satisfied, because displacement would not exist.

We shall therefore speak of *associated* and *complementary* differential equations in accordance with the following table:

TABLE I.

Type of boundary condition	Associated differential equations	Complementary differential equations
Stress (SBC)	Equilibrium (2.8)	Compatibility (2.6)
Displacement (DBC)	Compatibility (2.6)	Equilibrium (2.8)

This terminology will permit us to discuss both types of boundary condition with a single argument, which bifurcates only for details.

We note that the solution of an elastic problem involves the satisfaction of

(1) the associated differential equations,

(2) the complementary differential equations,

(3) the boundary conditions.

The following table sets forth schematically notation and terms to be used:

TABLE II.

Symbol for State	Name	Differential Equations Satisfied	Stress Boundary Conditions (SBC)		Displacement Boundary Conditions (DBC)	
			Diff. Equations Satisfied	Boundary Condns. Satisfied	Diff. Equations Satisfied	Boundary Condns. Satisfied
S	Natural state	Associated and Complementary	Equilibrium and Compatibility	$[T_{(n)}, T_{(t)}]$ $[T_{(n)}, u_{(t)}=0]$ $[T_{(t)}, u_{(n)}=0]$ $[u_{(n)}=0, u_{(t)}=0]$	Compatibility and Equilibrium	$[u_{(n)}, u_{(t)}]$ $[u_{(n)}, T_{(t)}=0]$ $[u_{(t)}, T_{(n)}=0]$ $[T_{(n)}=0, T_{(t)}=0]$
S^*	Completely associated state	Associated	Equilibrium	$T^*_{(n)}=T_{(n)}$ where $T_{(n)}$ assigned for S; $T^*_{(t)}=T_{(t)}$ where $T_{(t)}$ assigned for S.	Compatibility	$u^*_{(n)}=u_{(n)}$ where $u_{(n)}$ assigned for S; $u^*_{(t)}=u_{(t)}$ where $u_{(t)}$ assigned for S.
S'_p $(p=1, \cdots, m)$	Homogeneous associated states	Associated	Equilibrium	$T'_{(n)}=0$ where $T_{(n)}$ assigned for S; $T'_{(t)}=0$ where $T_{(t)}$ assigned for S.	Compatibility	$u'_{(n)}=0$ where $u_{(n)}$ assigned for S; $u'_{(t)}=0$ where $u_{(t)}$ assigned for S.
S''_q $(q=1, \cdots, n)$	Complementary states	Complementary	Compatibility	$u''_{(n)}=0$ where $u_{(n)}=0$ for S; $u''_{(t)}=0$ where $u_{(t)}=0$ for S.	Equilibrium	$T''_{(n)}=0$ where $T_{(n)}=0$ for S; $T''_{(t)}=0$ where $T_{(t)}=0$ for S.
I'_p $(p=1, \cdots, m)$	Orthonormal homogeneous associated states	← As for S'_p →				
I''_q $(q=1, \cdots, n)$	Orthonormal complementary states	← As for S''_q →				

We note that $\mathbf{S}$, the *natural state*, is the solution of the elastic problem with which we are concerned. It is to be regarded as unknown. All the other states are artificial states which we build up with a view to getting an approximation to $\mathbf{S}$; they are to be regarded as known.

In any particular problem, the boundary conditions will be either SBC or DBC. Hence, in any particular problem, we shall have occasion to consult the two columns under SBC, or the two columns under DBC, according to the type of boundary condition.

The table is self-explanatory, but it may assist the reader if we interpret it in the case where the stress T_i is given all over the surface. Here we have SBC, and the following interpretation:

$\mathbf{S}^*$ is any state which satisfies the equations of equilibrium (2.8), and has at each point of the surface the assigned value of T_i.

$\mathbf{S}_p'$ form a set of states, each satisfying the equations of equilibrium and homogeneous boundary conditions $T_i=0$.

$\mathbf{S}_q''$ form a set of states, each satisfying the equations of compatibility, but not subjected to any boundary conditions.

$\mathbf{I}_p'$ form an orthonormal set of states, each being a linear combination of the states $\mathbf{S}_p'$; each therefore satisfies the equations of equilibrium and makes $T_i=0$ on the surface.

$\mathbf{I}_q''$ form an orthonormal set of states, each being a linear combination of the states $\mathbf{S}_q''$; each therefore satisfies the equations of compatibility.

The procedure for forming an orthonormal set by linear combination of linearly independent states is a well known procedure. The general conditions for orthonormality read

$$\begin{aligned} \mathbf{I}_p'\cdot\mathbf{I}_r' &= \delta_{pr} \qquad (p, r = 1, \cdots, m) \\ \mathbf{I}_q''\cdot\mathbf{I}_s'' &= \delta_{qs} \qquad (q, s = 1, \cdots, n). \end{aligned} \tag{7.1}$$

These conditions imply that each vector is of unit magnitude, and that all vectors in each set are mutually orthogonal.

We shall always assume that the states $\mathbf{S}_p'$ are chosen linearly independent; otherwise the number of orthonormal vectors $\mathbf{I}_p'$ would be less than the number of vectors $\mathbf{S}_p'$. The same remarks apply to the states $\mathbf{S}_q''$.

The following statements are easily verified:

I. *Any linear combination of homogeneous associated states is itself a homogeneous associated state.*

II. *If any linear combination of homogeneous associated states is added to a completely associated state, the resulting state*

$$\mathbf{S}^* + \sum_{p=1}^{m} a_p'\mathbf{S}_p' \tag{7.2}$$

is a completely associated state.

III. *Any linear combination of complementary states*

$$\sum_{q=1}^{n} a_q''\mathbf{S}_q'' \tag{7.3}$$

is itself a complementary state.

8. The associated and complementary hyperplanes. If, in (7.2), we assign arbitrary values to the constants a_p', we get a linear subspace or hyperplane of m dimensions. We shall call this the *associated hyperplane*, and denote it by L_m^*. Similarly, we define the *complementary hyperplane* by (7.3), and denote it by L_n''. Fig. 5 shows these hyperplanes for the case $m=1$, $n=1$, when, of course, they are straight lines in function space.

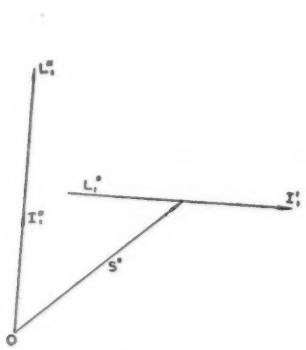

FIG. 5.

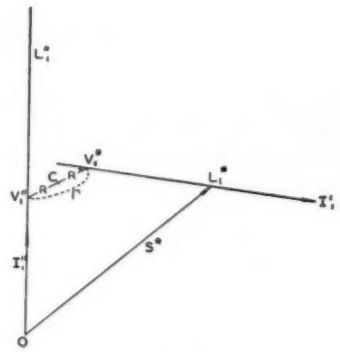

FIG. 6.

The associated hyperplane and the complementary hyperplane play important roles. If they possess a point of intersection, then the state corresponding to that intersection is the natural state **S**; for this state satisfies the associated differential equations, the complementary differential equations, and all the boundary conditions. The object of our approximations is to draw these hyperplanes together, and to locate their closest points, so that we may get a good approximation to **S**.

We shall now prove the following:

IV. *The associated hyperplane L_m^* and the complementary hyperplane L_n'' are mutually orthogonal.*

The hyperplanes are orthogonal if every vector lying wholly in one hyperplane is orthogonal to every vector lying wholly in the other. The general vector lying wholly in L_m^* is

$$\mathbf{S}' = \sum_{p=1}^{m} a_p' \mathbf{I}_p',$$

and the general vector lying wholly in L_n'' is

$$\mathbf{S}'' = \sum_{q=1}^{n} a_q'' \mathbf{I}_q''.$$

The necessary and sufficient condition for the orthogonality of the hyperplanes is

$$\mathbf{S}'.\mathbf{S}'' = 0. \tag{8.1}$$

To prove that this relation holds, we must consider SBC and DBC separately. For

SBC we have, since $\mathbf{S}'$ satisfies the equations of equilibrium and $\mathbf{S}''$ the equations of compatibility,

$$\mathbf{S}'.\mathbf{S}'' = (e''.E') = \int e''_{ij}E'_{ij}dv = \int u''_{i,j}E'_{ij}dv$$

$$= \int u''_i T'_i dS - \int u''_i E'_{ij,j}dv$$

$$= \int u''_{(n)}T'_{(n)}dS + \int u''_{(t)}T'_{(t)}dS, \tag{8.2}$$

the last integrand being the scalar product of vectors of displacement and stress lying in the tangent plane of the bounding surface. Let us now use the columns under SBC in Table II, and consider the integrals in the last line of (8.2) taken over the several regions into which the surface is divided. We see that

on	$[T_{(n)}, T_{(t)}]$	we have $T'_{(n)} = 0, T'_{(t)} = 0,$
on	$[T_{(n)}, u_{(t)} = 0]$	we have $T'_{(n)} = 0, u''_{(t)} = 0,$
on	$[T_{(t)}, u_{(n)} = 0]$	we have $T'_{(t)} = 0, u''_{(n)} = 0,$
on	$[u_{(n)} = 0, u_{(t)} = 0]$	we have $u''_{(n)} = 0, u''_{(t)} = 0.$

Hence the integrals in the last line of (8.2) vanish, and (8.1) is proved; this establishes the orthogonality of the two hyperplanes. The proof for DBC follows the same lines.

From (8.2) we have

$$\mathbf{I}'_p \cdot \mathbf{I}''_q = 0 \qquad (p = 1, 2, \cdots, m; q = 1, 2, \cdots, n). \tag{8.3}$$

9. Some basic relations. We shall now prove a number of results.

V. *For the natural state* $\mathbf{S}$ *and any completely associated state* $\mathbf{S}^*$, *the following relation holds*:

$$\mathbf{S}\cdot(\mathbf{S} - \mathbf{S}^*) = 0. \tag{9.1}$$

Hence the extremity of $\mathbf{S}$ *lies on a hypersphere having the vector* $\mathbf{S}^*$ *for diameter; the center is at* $\frac{1}{2}\mathbf{S}^*$ *and the radius is* $\frac{1}{2}S^*$.

To prove this, we separate the cases SBC and DBC. For SBC we have

$$\mathbf{S}\cdot\mathbf{S} = (u.T) = \int u_{(n)}T_{(n)}dS + \int u_{(t)}T_{(t)}dS,$$

$$\mathbf{S}\cdot\mathbf{S}^* = (u.T^*) = \int u_{(n)}T^*_{(n)}dS + \int u_{(t)}T^*_{(t)}dS. \tag{9.2}$$

By Table II,

on	$[T_{(n)}, T_{(t)}]$	we have $T^*_{(n)} = T_{(n)}, T^*_{(t)} = T_{(t)}$
on	$[T_{(n)}, u_{(t)} = 0]$	we have $T^*_{(n)} = T_{(n)}, u_{(t)} = 0,$
on	$[T_{(t)}, u_{(n)} = 0]$	we have $T^*_{(t)} = T_{(t)}, u_{(n)} = 0,$
on	$[u_{(n)} = 0, u_{(t)} = 0]$	we have $u_{(n)} = 0, u_{(t)} = 0.$

On using these values in (9.2), we see that $\mathbf{S}\cdot\mathbf{S}=\mathbf{S}\cdot\mathbf{S}^*$, and so (9.1) is proved for SBC. The proof for DBC is similar. Since (9.1) may be written

$$(\mathbf{S}-\tfrac{1}{2}\mathbf{S}^*)^2=\tfrac{1}{4}S^{*2}, \tag{9.3}$$

the truth of the statement about the hypersphere is obvious.

VI. *The natural state* $\mathbf{S}$ *is orthogonal to any homogeneous associated state* $\mathbf{S}'$:

$$\mathbf{S}\cdot\mathbf{S}'=0. \tag{9.4}$$

Hence $\mathbf{S}$ *is orthogonal to the associated hyperplane* L_m^* *and*

$$\mathbf{S}\cdot\mathbf{I}_p'=0 \qquad (p=1,2,\cdots,m). \tag{9.5}$$

Since any homogeneous associated state may be expressed as the difference between two completely associated states (cf. Table II), (9.4) follows if we write down relations of the form (9.1) for the two completely associated states and subtract one from the other.

VII. *The difference between the natural state* $\mathbf{S}$ *and any completely associated state* $\mathbf{S}^*$ *is orthogonal to any complementary state* $\mathbf{S}''$:

$$(\mathbf{S}-\mathbf{S}^*)\cdot\mathbf{S}''=0. \tag{9.6}$$

Hence this difference is orthogonal to the complementary hyperplane L_n'' *and*

$$\mathbf{S}\cdot\mathbf{I}_q''=\mathbf{S}^*\cdot\mathbf{I}_q'' \qquad (q=1,2,\cdots,n). \tag{9.7}$$

For SBC we have

$$\begin{aligned}\mathbf{S}\cdot\mathbf{S}''&=(e''\cdot E)=(u''\cdot T)=\int u''_{(n)}T_{(n)}dS+\int u''_{(t)}T_{(t)}dS,\\ \mathbf{S}^*\cdot\mathbf{S}''&=(e''\cdot E^*)=(u''\cdot T^*)=\int u''_{(n)}T^*_{(n)}dS+\int u''_{(t)}T^*_{(t)}dS.\end{aligned} \tag{9.8}$$

On reference to Table II, it is easy to see that these expressions are equal, and so (9.6) is established for SBC. The proof for DBC is similar.

VIII. *The difference between the natural state* $\mathbf{S}$ *and any completely associated state* $\mathbf{S}^*$ *is orthogonal to the difference between* $\mathbf{S}$ *and any complementary state* $\mathbf{S}''$:

$$(\mathbf{S}-\mathbf{S}^*)\cdot(\mathbf{S}-\mathbf{S}'')=0. \tag{9.9}$$

Thus the natural state lies on every hypersphere which has for diameter the line joining a point of the associated hyperplane to a point of the complementary hyperplane.

Equation (9.9) is an immediate consequence of (9.1) and (9.6). The statement about the hypersphere is verified by writing (9.9) in the form

$$[\mathbf{S}-\tfrac{1}{2}(\mathbf{S}^*+\mathbf{S}'')]^2=\tfrac{1}{4}(\mathbf{S}^*-\mathbf{S}'')^2. \tag{9.10}$$

10. The hypercircle. In (9.1) we located the natural state $\mathbf{S}$ on a hypersphere, and in (9.5), (9.7) we located it on $m+n$ hyperplanes. This means that the extremity of the natural vector $\mathbf{S}$ lies on a hypercircle Γ, which is the intersection of the hypersphere and the hyperplanes.

The center $\mathbf{C}$ of Γ may be found by starting from the center of the hypersphere

($\frac{1}{2}\mathbf{S}^*$), and proceeding through suitable distances in directions normal to the hyperplanes until we arrive at a point on all the hyperplanes. Thus, we write

$$\mathbf{C} = \tfrac{1}{2}\mathbf{S}^* + \sum_{p=1}^{m} a'_p \mathbf{I}'_p + \sum_{q=1}^{n} a''_q \mathbf{I}''_q, \tag{10.1}$$

where the coefficients are to be determined from the conditions

$$\mathbf{C}\cdot\mathbf{I}'_p = 0 \quad (p = 1, 2, \cdots, m), \qquad \mathbf{C}\cdot\mathbf{I}''_q = \mathbf{S}^*\cdot\mathbf{I}''_q \quad (q = 1, 2, \cdots, n). \tag{10.2}$$

By (7.1) and (8.3), we find

$$a'_p = -\tfrac{1}{2}\mathbf{S}^*\cdot\mathbf{I}'_p, \qquad a''_q = \tfrac{1}{2}\mathbf{S}^*\cdot\mathbf{I}''_q, \tag{10.3}$$

and so the center $\mathbf{C}$ of the hypercircle Γ is at

$$\mathbf{C} = \frac{1}{2}\left[\mathbf{S}^* - \sum_{p=1}^{m} \mathbf{I}'_p(\mathbf{S}^*\cdot\mathbf{I}'_p) + \sum_{q=1}^{n} \mathbf{I}''_q(\mathbf{S}^*\cdot\mathbf{I}''_q)\right]. \tag{10.4}$$

We note that the magnitude of $\mathbf{C}$ is given by

$$C^2 = \frac{1}{4}\left[S^{*2} - \sum_{p=1}^{m} (\mathbf{S}^*\cdot\mathbf{I}'_p)^2 + 3\sum_{q=1}^{n} (\mathbf{S}^*\cdot\mathbf{I}''_q)^2\right]. \tag{10.5}$$

We find the radius R of Γ from

$$\begin{aligned} R^2 &= (\mathbf{S} - \mathbf{C})^2 \\ &= \mathbf{S}\cdot(\mathbf{S} - \mathbf{S}^*) + \mathbf{S}\cdot(\mathbf{S}^* - 2\mathbf{C}) + C^2. \end{aligned} \tag{10.6}$$

The first term vanishes by (9.1), and we easily find from (10.4) and (10.5)

$$R^2 = \frac{1}{4}\left[S^{*2} - \sum_{p=1}^{m} (\mathbf{S}^*\cdot\mathbf{I}'_p)^2 - \sum_{q=1}^{n} (\mathbf{S}^*\cdot\mathbf{I}''_q)^2\right]. \tag{10.7}$$

To sum up:

IX. *The natural state* $\mathbf{S}$ *is located on a hypercircle* Γ *with center* $\mathbf{C}$ *as given by* (10.4) *and radius* R *as given by* (10.7). *Using* $\mathbf{X}$ *as a current position vector, we may write the equations of* Γ *in the form*

$$\begin{gathered} \mathbf{X}\cdot(\mathbf{X} - \mathbf{S}^*) = 0, \qquad \mathbf{X}\cdot\mathbf{I}'_p = 0 \quad (p = 1, 2, \cdots, m), \\ \mathbf{X}\cdot\mathbf{I}''_q = \mathbf{S}^*\cdot\mathbf{I}''_q \quad (q = 1, 2, \cdots, n). \end{gathered} \tag{10.8}$$

Let us now approach the hypercircle in a different way.

Points on the associated hyperplane L^*_m satisfy some of the conditions imposed on $\mathbf{S}$, and points on the complementary hyperplane L''_n satisfy others. If the two hyperplanes had a common point, then (as remarked earlier) that common point would be the natural state $\mathbf{S}$. Unless we are particularly lucky in choosing the states $\mathbf{S}^*$, $\mathbf{S}'_p$, $\mathbf{S}''_q$, this common point will not occur, and then it seems appropriate to seek those points, $\mathbf{V}^*_m$ on L^*_m and $\mathbf{V}''_n$ on L''_n such that the distance $|\mathbf{V}^*_m - \mathbf{V}''_n|$ is as small as possible. We may call these two points the *vertices* of the hyperplanes.

The general point on L^*_m is

$$\mathbf{U}^* = \mathbf{S}^* + \sum_{p=1}^{m} b'_p \mathbf{I}'_p, \tag{10.9}$$

and the general point on L_n'' is

$$\mathbf{U}'' = \sum_{q=1}^{n} b_q'' \mathbf{I}_q''. \tag{10.10}$$

The square of the distance between these two points is

$$(\mathbf{U}^* - \mathbf{U}'')^2 = \left(\mathbf{S}^* + \sum_{p=1}^{m} b_p' \mathbf{I}_p' - \sum_{q=1}^{n} b_q'' \mathbf{I}_q''\right)^2. \tag{10.11}$$

It is easily seen that to minimize this, we must choose

$$b_p' = -\mathbf{S}^* \cdot \mathbf{I}_p', \qquad b_q'' = \mathbf{S}^* \cdot \mathbf{I}_q''. \tag{10.12}$$

Substitution in (10.9) and (10.10) gives for the vertices of the associated and complementary hyperplanes respectively

$$\begin{aligned} \mathbf{V}_m^* &= \mathbf{S}^* - \sum_{p=1}^{m} \mathbf{I}_p' (\mathbf{S}^* \cdot \mathbf{I}_p') \\ \mathbf{V}_n'' &= \sum_{q=1}^{n} \mathbf{I}_q'' (\mathbf{S}^* \cdot \mathbf{I}_q''). \end{aligned} \tag{10.13}$$

We see from (10.4) that

$$\mathbf{C} = \tfrac{1}{2}(\mathbf{V}_m^* + \mathbf{V}_n''), \tag{10.14}$$

and from (10.7)

$$R^2 = \tfrac{1}{4}(\mathbf{V}_m^* - \mathbf{V}_n'')^2. \tag{10.15}$$

Thus the points $\mathbf{V}_m^*$ and $\mathbf{V}_n''$ are the extremities of a diameter of the hypercircle Γ.

Any chord of Γ may be represented by a vector $\mathbf{Y}$; it follows from (10.8) that every such chord satisfies $\mathbf{Y} \cdot \mathbf{I}_q'' = 0$ $(q = 1, 2, \cdots n)$, and hence $\mathbf{Y} \cdot \mathbf{V}_n'' = 0$.

We may sum up as follows:

X. *The vertices of the associated and complementary hyperplanes, as given in* (10.13), *are the extremities of a diameter of the hypercircle* Γ, *described in* IX. *The position vector* $\mathbf{V}_n''$ *of the vertex of the complementary hyperplane is orthogonal to every chord of* Γ (Fig. 6).

The above geometrical statement suggests the inequalities

$$V_n'' \leqq S \leqq V_m^*, \tag{10.16}$$

and a formal proof is easy. Any point $\mathbf{X}$ on Γ satisfies (cf. (10.8))

$$(\mathbf{X} - \mathbf{V}_m^*) \cdot (\mathbf{X} - \mathbf{V}_n'') = 0, \qquad \mathbf{X} \cdot \mathbf{V}_n'' = \mathbf{S}^* \cdot \mathbf{V}_n'', \tag{10.17}$$

and if we seek maxima and minima of X^2 subject to these conditions, it is easy to see that we must have

$$\mathbf{X} = a\mathbf{V}_m^* + b\mathbf{V}_n'', \tag{10.18}$$

where a and b are undetermined multipliers. When we substitute from this in (10.17), the second equation gives b as a linear expression in a, and then the first equation gives a quadratic for a. Thus there are just two solutions, and it is easy to verify from

(10.17) that they are $(a=1, b=0)$ and $(a=0, b=1)$, i.e. $\mathbf{X}=\mathbf{V}_m^*$ and $\mathbf{X}=\mathbf{V}_n''$. Moreover, by (10.13),

$$V_m^{*2} = S^{*2} - \sum_{p=1}^{m} (\mathbf{S}^* \cdot \mathbf{I}_p')^2,$$
$$V_n''^2 = \sum_{q=1}^{n} (\mathbf{S}^* \cdot \mathbf{I}_q'')^2, \tag{10.19}$$

and so, by (10.7),

$$V_m^{*2} - V_n''^2 = 4R^2 \geqq 0, \qquad V_m^* \geqq V_n''. \tag{10.20}$$

Thus V_m^* is the maximum of X and V_n'' the minimum, when $\mathbf{X}$ is any vector satisfying (10.17). Thus (10.16) is proved. To sum up:

XI. *The vertex of the associated hyperplane is further from the origin than the vertex of the complementary hyperplane, and the distance of the natural state from the origin is intermediate between them, so that*

$$V_n''^2 = \sum_{q=1}^{n} (\mathbf{S}^* \cdot \mathbf{I}_q'')^2 \leqq S^2 \leqq S^{*2} - \sum_{p=1}^{m} (\mathbf{S}^* \cdot \mathbf{I}_p'')^2 = V_m^{*2}. \tag{10.21}$$

These inequalities place lower and upper bounds on the strain energy of the natural state $\mathbf{S}$.

11. Approximations. If $\bar{\mathbf{S}}$ (a set of six stress components expressed as functions of position throughout the body) is suggested as an approximation to the natural state $\mathbf{S}$, the error, as defined in (4.14), is

$$\epsilon = |\bar{\mathbf{S}} - \mathbf{S}|, \tag{11.1}$$

and the squared error is

$$\epsilon^2 = (\bar{\mathbf{S}} - \mathbf{S})^2. \tag{11.2}$$

But we cannot in general calculate this error, because we do not know $\mathbf{S}$.

The hypercircle Γ comes to our assistance. The following is obvious:

XII. *If the center* $\mathbf{C}$ *of the hypercircle* Γ *as given by* (10.4), *namely*

$$\mathbf{C} = \frac{1}{2}\left[\mathbf{S}^* - \sum_{p=1}^{m} \mathbf{I}_p'(\mathbf{S}^* \cdot \mathbf{I}_p') + \sum_{q=1}^{n} \mathbf{I}_q''(\mathbf{S}^* \cdot \mathbf{I}_q'')\right] \tag{11.3}$$

is taken as an approximation to the natural state $\mathbf{S}$, *the error* ϵ *is precisely the radius* R *of* Γ, *so that by* (10.7)

$$4\epsilon^2 = 4R^2 = S^{*2} - \sum_{p=1}^{m} (\mathbf{S}^* \cdot \mathbf{I}_p')^2 - \sum_{q=1}^{n} (\mathbf{S}^* \cdot \mathbf{I}_q'')^2. \tag{11.4}$$

Since the distance between any two points on Γ cannot exceed the length of a diameter, we have the following result:

XIII. *If any state on the hypercircle* Γ *is taken as an approximation to* $\mathbf{S}$, *the error satisfies*

$$\epsilon \leqq 2R, \tag{11.5}$$

where R *is as in* (10.7). *The same inequality is satisfied if we take as an approximation*

either the vertex of the associated hyperplane,

$$\mathbf{V}_m^* = \mathbf{S}^* - \sum_{p=1}^{m} \mathbf{I}_p'(\mathbf{S}^* \cdot \mathbf{I}_p'), \tag{11.6}$$

or the vertex of the complementary hyperplane

$$\mathbf{V}_n'' = \sum_{q=1}^{n} \mathbf{I}_q''(\mathbf{S}^* \cdot \mathbf{I}_q''). \tag{11.7}$$

We note that since $\mathbf{C}$ does not satisfy the equations of compatibility, there is no corresponding displacement. Hence, if we are looking for an approximation for the displacement throughout the body, $\mathbf{V}_n''$ is the best approximation to take for SBC, and $\mathbf{V}_m^*$ the best for DBC.

We may add that if we take as an approximation

$$\bar{\mathbf{S}} = a\mathbf{V}_m^* + b\mathbf{V}_n'' \qquad (a + b = 1,\ a \geqq 0,\ b \geqq 0) \tag{11.8}$$

then ϵ satisfies

$$\begin{aligned} 2Rb \leqq \epsilon \leqq 2Ra \quad &\text{if} \quad b \leqq a, \\ 2Ra \leqq \epsilon \leqq 2Rb \quad &\text{if} \quad a \leqq b. \end{aligned} \tag{11.9}$$

Finally, we shall show how to use the hypercircle Γ to obtain lower and upper bounds for the error of any suggested approximation $\bar{\mathbf{S}}$. It is a question of finding the least and greatest distances of $\bar{\mathbf{S}}$ from the hypercircle with the equations (10.8). That means that we are to find the minimum and maximum values of

$$(\mathbf{X} - \bar{\mathbf{S}})^2$$

when the vector $\mathbf{X}$ is subject to the conditions (10.8).

It is clear that we must have

$$\mathbf{X} - \bar{\mathbf{S}} = a(2\mathbf{X} - \mathbf{S}^*) + \sum_{p=1}^{m} c_p'\mathbf{I}_p' + \sum_{q=1}^{n} c_q''\mathbf{I}_q'', \tag{11.10}$$

where the coefficients are undetermined multipliers. Equivalently,

$$(1 - 2a)\mathbf{X} = \bar{\mathbf{S}} - a\mathbf{S}^* + \sum_{p=1}^{m} c_p'\mathbf{I}_p' + \sum_{q=1}^{n} c_q''\mathbf{I}_q''. \tag{11.11}$$

Hence, by (10.8),

$$\begin{aligned} (1 - 2a)\mathbf{X} \cdot \mathbf{I}_p' &= c_p' + \bar{\mathbf{S}} \cdot \mathbf{I}_p' - a\mathbf{S}^* \cdot \mathbf{I}_p' = 0, \\ (1 - 2a)\mathbf{X} \cdot \mathbf{I}_q'' &= c_q'' + \bar{\mathbf{S}} \cdot \mathbf{I}_q'' - a\mathbf{S}^* \cdot \mathbf{I}_q'' = (1 - 2a)\mathbf{S}^* \cdot \mathbf{I}_q''. \end{aligned} \tag{11.12}$$

If we now substitute in (11.11) the values of c_p', c_q'' given by these equations, we get

$$(1 - 2a)\mathbf{X} = \bar{\mathbf{V}} - 2a\mathbf{C} \tag{11.13}$$

where $\mathbf{C}$ is the center of Γ, as in (11.3), and $\bar{\mathbf{V}}$ is defined as

$$\bar{\mathbf{V}} = \bar{\mathbf{S}} - \sum_{p=1}^{m} \mathbf{I}_p'(\bar{\mathbf{S}} \cdot \mathbf{I}_p') - \sum_{q=1}^{n} \mathbf{I}_q''[(\bar{\mathbf{S}} - \mathbf{S}^*) \cdot \mathbf{I}_q'']. \tag{11.14}$$

We may interpret (11.13) as follows: The points on Γ at minimax distances from $\mathbf{S}$ lie on the diameter of Γ which passes through $\bar{\mathbf{V}}$, and these points divide the line joining $\mathbf{C}$ to $\bar{\mathbf{V}}$ in the ratios $a:-\frac{1}{2}$. We have still to determine a.

The constant a is found by the first of (10.8). On substitution from (11.13), it is found that a satisfies the quadratic equation

$$4a^2\mathbf{C}\cdot(\mathbf{C}-\mathbf{S}^*)+2a[\bar{\mathbf{V}}\cdot(\mathbf{S}^*-\mathbf{C})+\mathbf{C}\,(\mathbf{S}^*-\bar{\mathbf{V}})]+\bar{\mathbf{V}}\cdot(\bar{\mathbf{V}}-\mathbf{S}^*)=0. \tag{11.15}$$

We may state the result as follows:

XIV. *The square error of any state* $\bar{\mathbf{S}}$ *lies between*

$$(\mathbf{X}_1-\bar{\mathbf{S}})^2 \quad \textit{and} \quad (\mathbf{X}_2-\bar{\mathbf{S}})^2,$$

where $\mathbf{X}_1$ *and* $\mathbf{X}_2$ *are the vectors given by* (11.13) *when the roots of* (11.15) *are substituted for* a.

Less precise but simpler bounds on the error of $\bar{\mathbf{S}}$ may be established by using, instead of the hypercircle Γ, the hypersphere with same center $\mathbf{C}$ and the same radius R. We first find whether $\bar{\mathbf{S}}$ lies outside or inside this hypersphere by computing

$$(\bar{\mathbf{S}}-\mathbf{C})^2-R^2. \tag{11.16}$$

If this is positive, $\bar{\mathbf{S}}$ is outside, if negative, inside. If $\bar{\mathbf{S}}$ is outside, we have

$$|\bar{\mathbf{S}}-\mathbf{C}|-R\leqq\epsilon\leqq|\bar{\mathbf{S}}-\mathbf{C}|+R. \tag{11.17}$$

If $\bar{\mathbf{S}}$ is inside, we have

$$R-|\bar{\mathbf{S}}-\mathbf{C}|\leqq\epsilon\leqq R+|\bar{\mathbf{S}}-\mathbf{C}|. \tag{11.18}$$

12. Example. As an illustration of the preceding relations, we consider the torsion of a prism of square cross section. We choose the centroid of one end section as the origin of coordinates and let the x_3 axis coincide with the axis of the prism. To avoid carrying along unessential constants, we consider the prism bounded by the planes $x_1=\pm1$, $x_2=\pm1$, $x_3=0$, $x_3=1$ and assume that the shear modulus G and the angle of twist ϑ equal unity. On the end sections of the prism we have $T_3=0$, and the tangential displacements u_α ($\alpha=1, 2$) are given, viz.

$$u_\alpha=0 \qquad \text{on} \quad x_3=0,$$
$$u_\alpha=-\epsilon_{\alpha\beta}x_\beta \quad \text{on} \quad x_3=1,$$

where $\epsilon_{11}=\epsilon_{22}=0$, $\epsilon_{12}=-\epsilon_{21}=1$. On the lateral faces of the prism $T_i=0$ ($i=1, 2, 3$). We thus have a case of DBC.

A suitable completely associated state is defined by

$$u_\alpha=-\epsilon_{\alpha\beta}x_\beta x_3, \qquad u_3=0, \tag{12.1}$$

or, with the shear modulus $G=1$,

$$F_{\alpha3}=-\epsilon_{\alpha\beta}x_\beta, \qquad E_{\alpha\beta}=E_{33}=0. \tag{S*}$$

We have

$$S^{*2}=\int_{-1}^{1}\int_{-1}^{1}E_{\alpha3}E_{\alpha3}dx_1dx_2=\int_{-1}^{1}\int_{-1}^{1}(x_1^2+x_2^2)dx_1dx_2=\frac{8}{3}\cdot \tag{12.2}$$

A homogeneous associated state is defined by

$$u_\alpha = 0, \qquad u_3 = \varphi(x_1, x_2). \tag{12.3}$$

The corresponding vector $\mathbf{S}'$ is

$$E_{\alpha 3} = \varphi_{,\alpha} \qquad E_{\alpha\beta} = E_{33} = 0, \tag{S'}$$

and we have

$$S'^2 = \int_{-1}^{1}\int_{-1}^{1} \varphi_{,\alpha}\varphi_{,\alpha} dx_1 dx_2. \tag{12.4}$$

As is easily seen, the "warping function" φ must assume values of equal magnitude but opposite signs in any two points which are symmetrically situated with respect to any one of the lines $x_1=0$, $x_2=0$, $x_1=x_2$, $x_1=-x_2$. A function of this type is

$$\varphi_p = x_1^{2p-1} x_2^{2p-1}(x_1^2 - x_2^2), \tag{12.5}$$

leading to the vector

$$\begin{aligned} E_{13} &= x_1^{2p-2} x_2^{2p-1}[(2p+1)x_1^2 + (2p-1)x_2^2], \\ E_{23} &= x_1^{2p-1} x_2^{2p-2}[(2p-1)x_1^2 + (2p+1)x_2^2]. \end{aligned} \tag{$\mathbf{S}_p'$}$$

We find

$$\mathbf{S}_p' \cdot \mathbf{S}_r' = \frac{2^5\{4[(p+r)^2 + 2pr - (p+r)] - 7\}}{[2(p+r)-3][2(p+r)-1][2(p+r)+1][2(p+r)+3]} \tag{12.6}$$

and

$$\mathbf{S}^* \cdot \mathbf{S}_p' = -\frac{2^4}{(2p+1)(2p+3)}. \tag{12.7}$$

A complementary state can be derived from a stress function $\psi(x_1, x_2)$ which vanishes along the contour of the cross section. The corresponding vector $\mathbf{S}''$ is

$$E_{\alpha 3} = \epsilon_{\alpha\beta}\psi_{,\beta}, \qquad E_{\alpha\beta} = E_{33} = 0, \tag{S''}$$

and we have

$$S''^2 = \int_{-1}^{1}\int_{-1}^{1} \psi_{,\alpha}\psi_{,\alpha} dx_1 dx_2. \tag{12.8}$$

As is easily seen, the stress function must be symmetric with respect to x_1 and x_2 and even in both these variables. Such a function which vanishes on the boundary of the cross section is

$$\psi_q = (x_1^{2q} - 1)(x_2^{2q} - 1), \tag{12.9}$$

leading to the vector

$$\begin{aligned} E_{13} &= 2q(x_1^{2q} - 1)x_2^{2q-1}, \\ E_{23} &= -2qx_1^{2q-1}(x_2^{2q} - 1). \end{aligned} \tag{$\mathbf{S}_q''$}$$

We find

$$\mathbf{S}_q''\cdot\mathbf{S}_s''=\frac{2^8q^2s^2(q+s+1)}{(2q+1)(2s+1)[2(q+s)-1][2(q+s)+1]} \tag{12.10}$$

and

$$\mathbf{S}^*\cdot\mathbf{S}_q''=\frac{2^5q^2}{(2q+1)^2}\cdot \tag{12.11}$$

We now proceed to determine the vertices $\mathbf{V}_m^*$ and $\mathbf{V}_m''$ in accordance with (10.13). We have

$$\mathbf{I}_1'(\mathbf{S}^*\cdot\mathbf{I}_1')=\mathbf{S}_1'(\mathbf{S}^*\cdot\mathbf{S}_1')/S_1'^2, \tag{12.12}$$

where

$$\mathbf{S}^*\cdot\mathbf{S}_1'=-\frac{2^4}{3\cdot5},\qquad S_1'^2=\frac{2^5\cdot3}{5\cdot7} \tag{12.13}$$

by Eqs. (12.7) and (12.6). Thus,

$$\mathbf{I}_1'(\mathbf{S}^*\cdot\mathbf{I}_1')=-\frac{7}{2\cdot3^2}\mathbf{S}_1',$$

and

$$\mathbf{V}_1^*=\mathbf{S}^*-\mathbf{I}_1'(\mathbf{S}^*\cdot\mathbf{I}_1')=\mathbf{S}^*+\frac{7}{2\cdot3^2}\mathbf{S}_1'. \tag{$\mathbf{V}_1^*$}$$

In a similar manner, we find

$$\mathbf{V}_1''=\mathbf{I}_1''(\mathbf{S}^*\cdot\mathbf{I}_1'')=\frac{5}{2^3}\mathbf{S}_1'' \tag{$\mathbf{V}_1''$}$$

To obtain $\mathbf{V}_2^*$ and $\mathbf{V}_2''$, we must first orthogonalize $\mathbf{S}_2'$ and $\mathbf{S}_2''$ with respect to $\mathbf{S}_1'$ and $\mathbf{S}_1''$, respectively. As is easily seen the vector

$$\tilde{\mathbf{S}}_2'=\mathbf{S}_2'-\mathbf{S}_1'(\mathbf{S}_1'\cdot\mathbf{S}_2')/S_1'^2 \tag{12.14}$$

is orthogonal to $\mathbf{S}_1'$. Using (12.6), we find

$$\tilde{\mathbf{S}}_2'=\mathbf{S}_2'-\frac{11}{3^3}\mathbf{S}_1' \tag{$\tilde{\mathbf{S}}_2'$}$$

and

$$\tilde{S}_2'^2=S_2'^2-(\mathbf{S}_1'\cdot\mathbf{S}_2')^2/S_1'^2=\frac{2^{12}}{3^5\cdot7\cdot11}\cdot \tag{12.15}$$

In a similar manner, a vector orthogonal to $\mathbf{S}_1''$ can be derived from $\mathbf{S}_2''$. One finds

$$\tilde{\mathbf{S}}_2''=\mathbf{S}_2''-\mathbf{S}_1''(\mathbf{S}_1''\cdot\mathbf{S}_2'')/S_1''^2=\mathbf{S}_2''-\frac{2^4\cdot3}{5\cdot7}\tilde{\mathbf{S}}_1'' \tag{$\tilde{\mathbf{S}}_2''$}$$

and

$$\tilde{S}_2''^2 = S_2''^2 - (\mathbf{S}_1'' \cdot \mathbf{S}_2'')^2 / S_1''^2 = \frac{2^{12} \cdot 31}{3^2 \cdot 5^3 \cdot 7^2}. \tag{12.16}$$

We now determine

$$\mathbf{I}_2' (\mathbf{S}^* \cdot \mathbf{I}_2') = \tilde{\mathbf{S}}_2' (\mathbf{S}^* \cdot \tilde{\mathbf{S}}_2') / \tilde{S}_2'^2.$$

With the use of Eqs. (12.6), (12.7) and (12.15), we find

$$\mathbf{I}_2' (\mathbf{S}^* \cdot \mathbf{I}_2') = -\frac{3 \cdot 11}{2^6 \cdot 5} \tilde{\mathbf{S}}_2'. \tag{12.17}$$

Similarly,

$$\mathbf{I}_2'' (\mathbf{S}^* \cdot \mathbf{I}_2'') = \frac{3 \cdot 5 \cdot 7}{2^5 \cdot 31} \tilde{\mathbf{S}}_2'' \tag{12.18}$$

Thus, by Eqs. (10.13),

$$\mathbf{V}_2^* = \mathbf{V}_1^* + \frac{3.11}{2^6 \cdot 5} \tilde{\mathbf{S}}_2' \tag{$\mathbf{V}_2^*$}$$

and

$$\mathbf{V}_2'' = \mathbf{V}_1'' + \frac{3 \cdot 5 \cdot 7}{2^5 \cdot 31} \tilde{\mathbf{S}}_2'' \tag{$\mathbf{V}_2''$}$$

In connection with the evaluation of (10.21), we need the values of $(\mathbf{S}^* \cdot \mathbf{I}_p')^2$ and $(\mathbf{S}^* \cdot \mathbf{I}_q'')^2$. One finds

$$(\mathbf{S}^* \cdot \mathbf{I}_1')^2 = (\mathbf{S}^* \cdot \mathbf{S}_1')^2 / S_1'^2 = \frac{2^3 \cdot 7}{3^3 \cdot 5} > 0.4148$$

$$(\mathbf{S}^* \cdot \mathbf{I}_2')^2 = \frac{11}{3^3 \cdot 5^2 \cdot 7} > 0.0023,$$

$$(\mathbf{S}^* \cdot \mathbf{I}_1'')^2 = \frac{2^2 \cdot 5}{3^2} > 2.2222,$$

$$(\mathbf{S}^* \cdot \mathbf{I}_2'')^2 = \frac{2^2}{5 \cdot 31} > 0.0258.$$

Now, S^2 here equals the product of the torque and the angle of twist. Since the latter has been taken as unity, S^2 represents the torque per unit angle of twist or the torsional stiffness. Equation (10.21) thus furnishes the following bounds for the torsional stiffness:*

a) for $m = n = 1$: $2.2222 < S^2 < \frac{8}{3} - 0.4148 < 2.2519$.

b) for $m = n = 2$: $2.2480 < S^2 < 2.2496$;

the exact value, to four decimal places, is $S^2 = 2.2492$.

* C. Weber [Z. angew. Math. Mech. **11**, 244–245 (1931)] used the variational principles of elasticity to obtain lower and upper bounds for the torsional stiffness of a square hollow prism. His basic formula corresponds essentially to the form which (10.21) assumes in the case of torsion for $m = n = 1$.

Next, let us consider the maximum shearing stress, i.e. the stress component E_{23} at the point $x_1=1$, $x_2=0$. For the various states considered above the value of this stress component is given in the following table:

State	$\mathbf{S}^*$	$\mathbf{S}_1'$	$\mathbf{S}_2'$	$\mathbf{S}_1''$	$\mathbf{S}_2''$	$\tilde{\mathbf{S}}_2'$	$\tilde{\mathbf{S}}_2''$
$E_{23}(1,0)$	1	1	0	2	4	$-11/3^3$	$2^2\cdot 11/(5\cdot 7)$

Accordingly, the stress values corresponding to the vertices of the associated and complementary hyperplanes are:

$$1+\frac{7}{2\cdot 3^2}=\frac{25}{18}=1.389 \quad \text{for} \quad \mathbf{V}_1^*,$$

$$\frac{25}{18}-\frac{3\cdot 11}{2^6\cdot 5}\cdot\frac{11}{3^3}=\frac{431}{320}=1.347 \quad \text{for} \quad \mathbf{V}_2^*,$$

$$\frac{5}{2^3}\cdot 2=\frac{5}{4}=1.250 \quad \text{for} \quad \mathbf{V}_1'' \quad \text{and}$$

$$\frac{5}{4}+\frac{3\cdot 5\cdot 7}{2^5\cdot 31}\cdot\frac{2^2\cdot 11}{5\cdot 7}=\frac{343}{248}=1.383 \quad \text{for} \quad \mathbf{V}_2''.$$

The stress values corresponding to

$$\mathbf{C}_1=\tfrac{1}{2}(\mathbf{V}_1^*+\mathbf{V}_1'')$$

and

$$\mathbf{C}_2=\tfrac{1}{2}(\mathbf{V}_2^*+\mathbf{V}_2'')$$

are therefore 1.319 and 1.365, respectively; the exact value, to three decimal places, is 1.352.

13. Comparison with other work. S. Bergman* has given an exact solution of the general boundary value problem for an isotropic elastic body, the surface stress or the surface displacement being assigned. In the vector notation of the present paper, his solution may be written

$$\mathbf{S}=\sum_{p=1}^{\infty}\mathbf{Z}_p(\mathbf{Z}_p\cdot\mathbf{S}),$$

where $\mathbf{Z}_p$ form an infinite orthonormal sequence of states, each satisfying the equations of equilibrium and compatibility. These states are determined explicitly. The scalar product $\mathbf{Z}_p\cdot\mathbf{S}$ can be computed when either stress or displacement is given all over the bounding surface. A point of Bergman's function space is a set of three functions (components of displacement), whereas a point of our function space is a set of six functions (components of stress), since we relax the equations of compatibility. But a more important difference between Bergman's work and the present theory lies in the fact that he was primarily concerned with an exact solution, whereas we are interested in comparatively simple approximate solutions in which the error can be easily computed.

It is interesting to discuss, along the lines of the present paper, the results given in

* Mathematische Annalen, **98**, 248–263 (1928).

Courant-Hilbert, "Methoden der mathematischen Physik," vol. 1, J. Springer, Berlin, 1931, pp. 228–230. We shall refer to this work as CH. We shall show how their results can be obtained without recourse to the calculus of variations. Then we shall strengthen these results, and express them as a single inequality. Finally, for the boundary conditions considered in CH, we shall show that the natural state lies on a certain hypersphere.

The boundary conditions used by CH are different from those of the present paper. In their simpler type of boundary condition, they divide the boundary into two parts, Γ_1 and Γ_2, with T_i assigned on Γ_1, and u_i assigned on Γ_2. Body forces are included in CH, but we shall here omit them.

Unless either Γ_1 or Γ_2 disappears, these boundary conditions are neither SBC nor DBC. We shall therefore abandon our previous notation, and use the following symbols:

$\mathbf{S}$ = natural state.

$\mathbf{S}'$ = a state satisfying the equations of equilibrium and making $T_i' = T_i$ on Γ_1.

$\mathbf{S}''$ = a state satisfying the equations of compatibility and making $u_i'' = u_i$ on Γ_2.

Now

$$\int (e_{ij} - e_{ij}')(E_{ij} - E_{ij}')dv \geqq 0, \tag{13.1}$$

or

$$S^2 + S'^2 - 2(e \cdot E') \geqq 0, \tag{13.2}$$

or, in an obvious extension of the notation of (2.10),

$$S^2 + S'^2 \geqq 2(u \cdot T')_1 + 2(u \cdot T')_2, \tag{13.3}$$

where the subscripts 1 and 2 refer to integration over Γ_1 and Γ_2 respectively. Also

$$S^2 = (u \cdot T)_1 + (u \cdot T)_2. \tag{13.4}$$

Subtracting twice (13.4) from (13.3) and using the fact that $T_i' = T_i$ on Γ_1, we get

$$S'^2 - S^2 \geqq 2(u \cdot T')_2 - 2(u \cdot T)_2, \tag{13.5}$$

or

$$\tfrac{1}{2}S^2 - (u \cdot T)_2 \leqq \tfrac{1}{2}S'^2 - (u \cdot T')_2. \tag{13.6}$$

This is the same as the principle II of CH. By a similar argument, we obtain

$$\tfrac{1}{2}S^2 - (u \cdot T)_1 \leqq \tfrac{1}{2}S''^2 - (u'' \cdot T)_1, \tag{13.7}$$

which is the same as the principle I of CH.

If we add (13.6) and (13.7), and use (13.4), we get

$$\tfrac{1}{2}S'^2 + \tfrac{1}{2}S''^2 - (u \cdot T')_2 - (u'' \cdot T)_1 \geqq 0, \tag{13.8}$$

a result which is of no particular interest.

Now let us strengthen the preceding results, using (3.9), which yields

$$\mathbf{S} \cdot \mathbf{S}' \leqq SS', \qquad \mathbf{S} \cdot \mathbf{S}'' \leqq SS''. \tag{13.9}$$

The first of these gives, instead of (13.3), the stronger inequality

$$SS' \geqq (u \cdot T')_1 + (u \cdot T')_2 = (u \cdot T)_1 + (u \cdot T')_2. \tag{13.10}$$

Subtracting (13.4), we get

$$SS' - S^2 \geqq (u \cdot T')_2 - (u \cdot T)_2, \tag{13.11}$$

or, instead of (13.6),

$$S^2 - (u \cdot T)_2 \leqq SS' - (u \cdot T')_2. \tag{13.12}$$

Similarly, instead of (13.7), we get

$$S^2 - (u \cdot T)_1 \leqq SS'' - (u'' \cdot T)_1. \tag{13.13}$$

When we added (13.6) and (13.7) we got the trivial inequality (13.8), which tells us nothing about the natural state. But if we add (13.12) and (13.13), and use (13.4), we get

$$S^2 \leqq S(S' + S'') - (u \cdot T')_2 - (u'' \cdot T)_1, \tag{13.14}$$

or

$$[S - \tfrac{1}{2}(S' + S'')]^2 \leqq \tfrac{1}{4}(S' + S'')^2 - (u \cdot T')_2 - (u'' \cdot T)_1. \tag{13.15}$$

This single inequality bounds the strain energy of the natural state below and above, for all quantities except S are calculable, and we have

$$\tfrac{1}{2}(S' + S'') - R \leqq S \leqq \tfrac{1}{2}(S' + S'') + R, \tag{13.16}$$

where $R > 0$ and

$$R^2 = \tfrac{1}{4}(S' + S'')^2 - (u \cdot T')_2 - (u'' \cdot T)_1. \tag{13.17}$$

But we can do better than this. We have

$$\begin{aligned} S^2 &= (u \cdot T)_1 + (u \cdot T)_2, \\ \mathbf{S} \cdot \mathbf{S}' &= (u \cdot T')_1 + (u \cdot T')_2, \\ \mathbf{S} \cdot \mathbf{S}'' &= (u'' \cdot T)_1 + (u'' \cdot T)_2. \end{aligned} \tag{13.18}$$

Since $(u \cdot T)_1 = (u \cdot T')_1$, $(u \cdot T)_2 = (u'' \cdot T)_2$, we get

$$S^2 - \mathbf{S} \cdot (\mathbf{S}' + \mathbf{S}'') = -(u \cdot T')_2 - (u'' \cdot T)_1, \tag{13.19}$$

or

$$[\mathbf{S} - \tfrac{1}{2}(\mathbf{S}' + \mathbf{S}'')]^2 = R^2, \tag{13.20}$$

where $R^2 = \tfrac{1}{4}(\mathbf{S}' + \mathbf{S}'')^2 - (u \cdot T')_2 - (u'' \cdot T)_1$. *Thus the natural state lies on a hypersphere with center at* $\tfrac{1}{2}(\mathbf{S}' + \mathbf{S}'')$ *and radius* R.

As remarked in a footnote to CH, their results apply to more general boundary conditions. These more general boundary conditions include our SBC and DBC as special cases, and may be described as follows: The boundary surface may be divided into four regions:

$$\begin{aligned} &\Gamma_{11}: \quad T_{(n)},\ T_{(t)} \text{ assigned.} \\ &\Gamma_{12}: \quad T_{(n)},\ u_{(t)} \text{ assigned.} \\ &\Gamma_{21}: \quad u_{(n)},\ T_{(t)} \text{ assigned.} \\ &\Gamma_{22}: \quad u_{(n)},\ u_{(t)} \text{ assigned.} \end{aligned} \tag{13.21}$$

We define

$\mathbf{S}$ = natural state.

$\mathbf{S}'$ = state satisfying equations of equilibrium and all boundary conditions on $T_{(n)}$ and $T_{(t)}$.

$\mathbf{S}''$ = state satisfying equations of compatibility and all boundary conditions on $u_{(n)}$ and $u_{(t)}$.

Then

$$S^2 = (u_{(n)}T_{(n)})_{11} + (u_{(t)}T_{(t)})_{11} + (u_{(n)}T_{(n)})_{12} + (u_{(t)}T_{(t)})_{12} + (u_{(n)}T_{(n)})_{21} + (u_{(t)}T_{(t)})_{21} + (u_{(n)}T_{(n)})_{22} + (u_{(t)}T_{(t)})_{22}, \tag{13.22}$$

$$\mathbf{S}\cdot\mathbf{S}' = \text{same expression as in (13.22) but with } T \text{ changed to } T', \tag{13.23}$$

$$\mathbf{S}\cdot\mathbf{S}'' = \text{same expression as in (13.22) but with } u \text{ changed to } u''. \tag{13.24}$$

Now we have

$$(u_{(n)}T'_{(n)})_{11} = (u_{(n)}T_{(n)})_{11}, \qquad (u_{(t)}T'_{(t)})_{11} = (u_{(t)}T_{(t)})_{11}$$
$$(u_{(n)}T'_{(n)})_{12} = (u_{(n)}T_{(n)})_{12}, \qquad (u_{(t)}T'_{(t)})_{21} = (u_{(t)}T_{(t)})_{21}$$
$$(u''_{(n)}T_{(n)})_{21} = (u_{(n)}T_{(n)})_{21}, \qquad (u''_{(t)}T_{(t)})_{12} = (u_{(t)}T_{(t)})_{12}$$
$$(u''_{(n)}T_{(n)})_{22} = (u_{(n)}T_{(n)})_{22}, \qquad (u''_{(t)}T_{(t)})_{22} = (u_{(t)}T_{(t)})_{22}.$$

So we obtain from (13.22), (13.23), (13.24),

$$S^2 - \mathbf{S}\cdot(\mathbf{S}' + \mathbf{S}'') = -(u''\cdot T'')_{11} - (u\cdot T')_{22} - (u_{(n)}T'_{(n)})_{21} - (u_{(t)}T'_{(t)})_{12} - (u''_{(n)}T_{(n)})_{12} - (u''_{(t)}T_{(t)})_{21}. \tag{13.26}$$

The right hand side is computable. The result may be written

$$[\mathbf{S} - \tfrac{1}{2}(\mathbf{S}' + \mathbf{S}'')]^2 = R^2, \tag{13.27}$$

where

$$R^2 = \tfrac{1}{4}(\mathbf{S}' + \mathbf{S}'')^2 - (u''\cdot T'')_{11} - (u\cdot T')_{22} - (u_{(n)}T'_{(n)})_{21} - (u_{(t)}T'_{(t)})_{12} - (u_{(n)}T''_{(n)})_{12} - (u_{(t)}T''_{(t)})_{22}. \tag{13.28}$$

We may sum up as follows:

XIV. *With the boundary conditions as described in* (13.21), *the natural state lies on a hypersphere with center at*

$$\mathbf{C} = \tfrac{1}{2}(\mathbf{S}' + \mathbf{S}'') \tag{13.29}$$

and radius R as given by (13.28), *where* $\mathbf{S}'$ *is any state satisfying the equations of equilibrium and all boundary conditions on stress, and* $\mathbf{S}''$ *is any state satisfying the equations of compatibility and all boundary conditions on displacement.*

Added November 18, 1946: The authors are indebted to Professor A. Weinstein for showing a connection between the present method and the method given by E. Trefftz.* Trefftz dealt with the Dirichlet problem in a plane (to determine a harmonic function with prescribed boundary values). The Ritz method gives an upper bound to the Dirichlet integral of the solution u, namely, the Dirichlet integral of a function v satisfying the boundary conditions. Trefftz supplied a lower bound by

* Proc. 2nd International Congress for Applied Mechanics, Zurich, 1926, pp. 131–137.

taking a linear combination w of harmonic functions, and choosing the coefficients so that the Dirichlet integral of the difference $(u-w)$ is a minimum. It should be pointed out that, although the torsion problem is reducible to the Dirichlet problem, the torsional rigidity is not given by the Dirichlet integral of a harmonic function. Hence, in applying his method to the estimation of torsional rigidity, Trefftz had to proceed in an indirect way.

Let us now consider the Dirichlet problem in the notation of the present paper. Let a vector in function space correspond to a function with continuous first derivatives in a plane region R, bounded by a curve B. Let $\mathbf{S}^*$ be a *completely associated vector*; this means a function u^* satisfying the boundary conditions. Let $\mathbf{S}''$ be a *complementary vector;* this means a *harmonic function* u''. Let $\mathbf{S}$ be the solution-function u. Let the scalar product of two vectors be defined by the Dirichlet integral

$$\mathbf{S}_1 \cdot \mathbf{S}_2 = \int\int_R \left(\frac{\partial u_1}{\partial x} \frac{\partial u_2}{\partial x} + \frac{\partial u_1}{\partial y} \frac{\partial u_2}{\partial y} \right) dxdy.$$

Then

$$\mathbf{S}^* \cdot \mathbf{S}'' = \int\int_R \left(\frac{\partial u^*}{\partial x} \frac{\partial u''}{\partial x} + \frac{\partial u^*}{\partial y} \frac{\partial u''}{\partial y} \right) dxdy = \int_B u^* \frac{\partial u''}{\partial n} ds,$$

since u'' is harmonic. Since $\mathbf{S}$ is itself a completely associated vector, we may write $\mathbf{S}$ for $\mathbf{S}^*$ and u for u^*. But $u^* = u$ on B, and so we have

$$\mathbf{S}^* \cdot \mathbf{S}'' = \mathbf{S} \cdot \mathbf{S}'' \quad \text{or} \quad \mathbf{S}'' \cdot (\mathbf{S}^* - \mathbf{S}) = 0. \tag{1}$$

But $\mathbf{S}$ is also a complementary vector, and so we may write $\mathbf{S}$ instead of $\mathbf{S}''$. This gives

$$\mathbf{S}^* \cdot \mathbf{S} = \mathbf{S}^2 \quad \text{or} \quad \mathbf{S} \cdot (\mathbf{S}^* - \mathbf{S}) = 0. \tag{2}$$

This last equation is the Ritz equation (4) of Trefftz' paper.

The equation (1) is not that of Trefftz. However, it tells us that $\mathbf{S}^* - \mathbf{S}$ is perpendicular to $\mathbf{S}''$. From elementary solid geometry it follows that, if we drop perpendiculars from the extremities of $\mathbf{S}^*$ and $\mathbf{S}$ on the line of $\mathbf{S}''$, these perpendiculars have a common foot, say with position vector $\mathbf{T}'' = k\mathbf{S}''$. We have then

$$\mathbf{T}'' \cdot (\mathbf{S} - \mathbf{T}'') = 0. \tag{3}$$

This is identical with Trefftz' equation (10), in which the right side is actually zero. Obviously, $\mathbf{T}''$ is that vector of form $k\mathbf{S}''$ which minimizes $(\mathbf{S} - \mathbf{T}'')^2$; thus $\mathbf{T}''$ corresponds to that harmonic function of form ku'' which minimizes the Dirichlet integral of $u - ku''$, where u is the solution.

To apply the above ideas to the torsion problem, we recall that the torsional rigidity (for $G = 1$) is

$$\int\int_R \left(x^2 + y^2 - x \frac{\partial u}{\partial x} - y \frac{\partial u}{\partial y} \right) dxdy,$$

where u is a harmonic function, taking the value $\frac{1}{2}(x^2+y^2)$ on the boundary. If we write $\mathbf{S}_0$ for the vector $\frac{1}{2}(x^2+y^2)$, then the torsional rigidity is

$$\mathbf{S}_0^2 - \mathbf{S}_0 \cdot \mathbf{S}. \tag{4}$$

By virtue of (1) and (2) we can build up the theory of the hypercircle, as we did in the general elastic problem. We can introduce orthonormal homogeneous associated vectors $\mathbf{I}_p'$ $(p=1, \cdots, m)$ with $u'=0$ on B, and orthonormal complementary vectors $\mathbf{I}_q''$ $(q=1, \cdots, n)$, each harmonic. The center $\mathbf{C}$ of the hypercircle is then given by (10.4) and its radius by (10.7). We have

$$\begin{aligned} (\mathbf{S}-\mathbf{C})\cdot\mathbf{I}_p' &= 0 \qquad (p=1, \cdots, m) \\ (\mathbf{S}-\mathbf{C})\cdot\mathbf{I}_q'' &= 0 \qquad (q=1, \cdots, n). \end{aligned} \tag{5}$$

Thus we may write

$$\mathbf{S} = \mathbf{C} + R\mathbf{J}, \tag{6}$$

where $\mathbf{J}$ is a unit vector satisfying

$$\begin{aligned} \mathbf{J}\cdot\mathbf{I}_p' &= 0 \qquad (p=1, \cdots, m) \\ \mathbf{J}\cdot\mathbf{I}_q'' &= 0 \qquad (q=1, \cdots, n). \end{aligned} \tag{7}$$

Thus

$$\mathbf{S}_0\cdot\mathbf{S} = \mathbf{S}_0\cdot\mathbf{C} + R\mathbf{S}_0\cdot\mathbf{J}. \tag{8}$$

The maximum and minimum of this expression, as $\mathbf{S}$ ranges over the hypercircle, are to be found from

$$\mathbf{S}_0 = \sum_{p=1}^{m} c_p'\mathbf{I}_p' + \sum_{q=1}^{n} c_q''\mathbf{I}_q'' + c\mathbf{J}, \tag{9}$$

where the c's are undetermined scalars. Since $\mathbf{I}_p'\cdot\mathbf{I}_q''=0$, we easily find

$$\begin{aligned} c_p' &= \mathbf{S}_0\cdot\mathbf{I}_p' \qquad (p=1, \cdots, m) \\ c_q'' &= \mathbf{S}_0\cdot\mathbf{I}_q'' \qquad (q=1, \cdots, n) \\ c^2 &= \mathbf{S}_0^2 - \sum_{p=1}^{m} (\mathbf{S}_0\cdot\mathbf{I}_p')^2 - \sum_{q=1}^{n} (\mathbf{S}_0\cdot\mathbf{I}_q'')^2 \\ \mathbf{S}_0\cdot\mathbf{J} &= c. \end{aligned} \tag{10}$$

Hence by (8)

$$\mathbf{S}_0\cdot\mathbf{C} - RA \leqq \mathbf{S}_0\mathbf{S} \leqq \mathbf{S}_0\cdot\mathbf{C} + RA, \tag{11}$$

where

$$A = \left[\mathbf{S}_0^2 - \sum_{p=1}^{m} (\mathbf{S}_0\cdot\mathbf{I}_p')^2 - \sum_{q=1}^{n} (\mathbf{S}_0\cdot\mathbf{I}_q'')\right]^{1/2}.$$

So we have upper and lower bounds for the torsional rigidity, as given by (4).

THE FLOW OF AN IDEAL INCOMPRESSIBLE FLUID ABOUT A LENS*

BY

MAX SHIFFMAN (*New York University*) AND D. C. SPENCER (*Stanford University*)

1. Introduction. The irrotational flow produced in an ideal incompressible fluid by the motion of an immersed object is described by a velocity potential function satisfying certain boundary conditions. For three-dimensional flows the velocity potential has been determined explicitly only for a very limited category of geometrically prescribed objects. In the present paper, we shall add to this list by constructing the flow about a lenticular object formed by two intersecting spheres, typical examples of which are illustrated in Figs. 1a–c. This lens flow is closely related to the well known case of two completely separated spheres, which might from a general point of view be considered as two spheres having an imaginary circle of intersection. In fact, it is this relationship which forms the point of departure of the present work and which will be clarified herein.

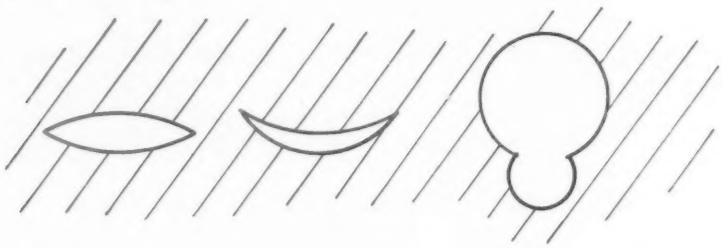

FIG. 1a FIG. 1b FIG. 1c

We shall consider, then, a lens formed by two intersecting spheres submerged in an ideal incompressible fluid and moving at unit velocity in the direction of its axis of symmetry. The velocity potential for the analogous two-dimensional problem could be easily determined by the use of conformal mapping but this is of no avail for three dimensions. Rather, we shall continue the potential function Φ analytically as far as possible across the two faces of the lens, and determine its singularities. The extended function Φ is then defined over a multi-sheeted—in general infinite-sheeted—Riemann space with the sharp rim of the lens, the circle of intersection of the two spheres, as the branch line around which the Riemann space winds. By the nature of the analytic consideration, the singularities are located not only at the inverted points but also at the points at infinity in the various sheets. It is then necessary to construct fundamental potential functions defined on the multisheeted Riemann space and having prescribed singularities at given points. For this construction we shall use the same type of procedure as was introduced by A. Sommerfeld in his classical work on the theory of diffraction, where a similar question was considered for reflection across plane surfaces.

* Received Sept. 27, 1946. This paper is essentially an excerpt from a report by the authors entitled "The Force of Impact on a Sphere Striking a Water Surface," AMP Report 42.1R, Feb. 1945, Appl. Math. Group at New York University, prepared for the Office of Scientific Research and Development. The application to impact will be published separately.

2. Procedure for separated spheres. Let Φ be the velocity potential (we use the convention that velocity $= -$grad Φ). It satisfies the Laplace equation

$$\Phi_{xx} + \Phi_{yy} + \Phi_{zz} = 0,$$

vanishes and is regular at ∞, and satisfies the following boundary condition: on the surface of the lens, the negative of the normal derivative of Φ is equal to the component of the velocity of the lens along its normal, or

$$-\frac{\partial \Phi}{\partial n} = \cos \lambda.$$

To see how to construct the potential function Φ, let us first consider the case in which the two spheres forming the lens do not intersect and are thus completely separated (Fig. 2).

A classical procedure for finding the flow due to the vertical motion of the two spheres is by the method of images (see [1], p. 131 or [2], p. 420).* The flow produced by a single sphere alone moving in an infinite body of fluid is that created by a dipole of suitable strength placed at the center of the sphere. Thus, dipoles situated at O and O' in the figure would individually give the correct flow if the influence of the other sphere could be neglected. But the dipole at O' destroys the correct boundary condition on the sphere K, and likewise for the dipole at O and the sphere K'. To remedy this, introduce the image of the dipole at O' relative to K (and likewise for O and K'), namely, a dipole of suitable strength situated at the inverse point Q_1 (and at the inverse point Q_1'). But again the dipole at Q_1' violates the boundary condition on K, etc. Indefinite continuation of this process leads to a convergent expression for the potential function which satisfies the boundary conditions on both spheres.

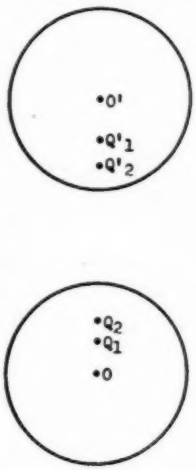

FIG. 2

We shall apply a generalization of the above process to the lenticular case, when the spheres intersect. The process as it stands leads to a serious difficulty, namely that some of the successive images of the dipoles lie directly in the fluid, thus creating singularities inside the fluid which violate the requirements of the problem. For examples, in the particular case of Fig. 1a, the very first dipoles at the centers of the respective spheres already lie in the fluid. The difficulty may be overcome by using multivalued dipoles, the images of which lie in the various replicas ("sheets") of a Riemann space and do not insert singularities in that portion of the space in which we are interested. However, there is no difficulty in the special cases when the exterior angle between the intersecting spheres has the value π/m where m is an integer. This is discussed in App. A. It should of course be mentioned that when the exterior angle between the intersecting spheres is greater than π, as in Figs. 1a, b, the velocity of the fluid is infinite on the sharp rim of the lens.

3. Dipoles. We select the toroidal coordinate system (σ, ψ, θ) as most appropriate for our domain. These are formed from bipolar coordinates in the plane by rotation about an axis, and are defined by

* Numbers in brackets refer to the bibliography at the end of the paper.

$$x = r\cos\varphi, \qquad y = r\sin\varphi, \qquad r + iz = ai\cot\tfrac{1}{2}(\psi + i\sigma)$$

where x, y, z are rectangular coordinates, $r^2 = x^2 + y^2$, and r, z can be expressed separately in terms of σ, ψ by

$$r = a\frac{\sinh\sigma}{s - \tau}, \qquad z = a\frac{\sin\psi}{s - \tau}$$

where

$$s = \cosh\sigma, \qquad \tau = \cos\psi.$$

(See [2], p. 165.) Because of rotational symmetry, the angle θ will never enter. In the r, z-plane, the curve ψ = constant is a circular arc connecting the points $+a$ and $-a$ on the r-axis, while σ = constant is a circle with these two points as mutually inverse points. By letting ψ vary from $-\infty$ to $+\infty$ without any identification, there results an infinite-fold Riemann space with the circle $r = a$, $z = 0$ as branch line.

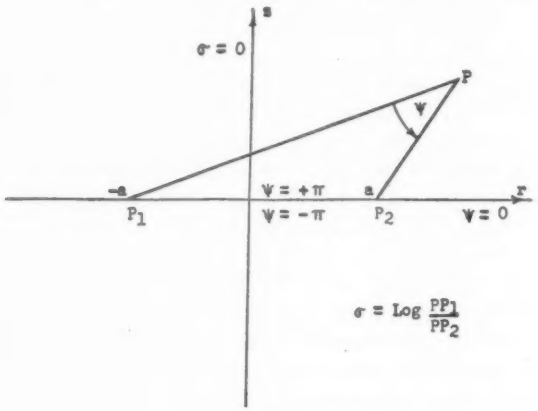

FIG. 3

The inverse distance between a point on the z-axis $\sigma = 0$ at $\psi = \psi_0$ and an arbitrary point (σ, ψ, θ) is the absolute value of the quantity

$$p(\psi_0) = \frac{\sin(\psi_0/2)}{a}\sqrt{\frac{s - \tau}{s - \cos(\psi - \psi_0)}}.$$

This represents a pole of strength ± 1 at the point $\sigma = 0$, $\psi = \psi_0$. We shall consider a dipole $q(\psi_0)$ of strength $a^3|\csc^3(\psi_0/2)|/2$ situated at the same point, and pointing downwards, namely

$$\begin{aligned} q(\psi_0) &= -\frac{a^3}{2}\csc^3(\psi_0/2)\frac{\partial}{\partial z}p(\psi_0) \\ &= \frac{a}{2}\left\{\frac{(s - \tau)^{1/2}\sin(\psi - \psi_0)}{[s - \cos(\psi - \psi_0)]^{3/2}} + \frac{(s - \tau)^{1/2}\cot(\psi_0/2)}{[s - \cos(\psi - \psi_0)]^{1/2}}\right\}. \end{aligned} \tag{3.1}$$

The Stokes' stream function $t(\psi_0)$ corresponding to $q(\psi_0)$ is (see [2], chapter XV, especially page 414):

$$t(\psi_0) = \tfrac{1}{2}a^3 \csc^3(\psi_0/2) r^2 p^3(\psi_0) = \frac{a^2 \sinh^2 \sigma}{2(s-\tau)^{1/2}[s-\cos(\psi-\psi_0)]^{3/2}}. \tag{3.2}$$

From this it follows that the potential function

$$q(\psi_0) - q(\psi_0')$$

has vanishing normal derivative on the spherical cap $\psi = c$ if

$$\psi_0 + \psi_0' = 2c.$$

For then $t(\psi_0) - t(\psi_0')$ vanishes on $\psi = c$.

4. Multi-valued dipoles. The quantities $q(\xi)$, $t(\xi)$ remain as potential and stream functions respectively of a flow if ψ_0 is replaced by a complex number ξ. Considering σ, ψ temporarily as fixed, $q(\xi)$ and $t(\xi)$ as functions of ξ in the complex ξ-plane are multi-valued and have branch points at the zeros of $s - \cos(\psi - \xi)$, that is at

$$\xi = \psi \pm i\sigma + 2k\pi, \qquad k = 0, \pm 1, \pm 2, \cdots.$$

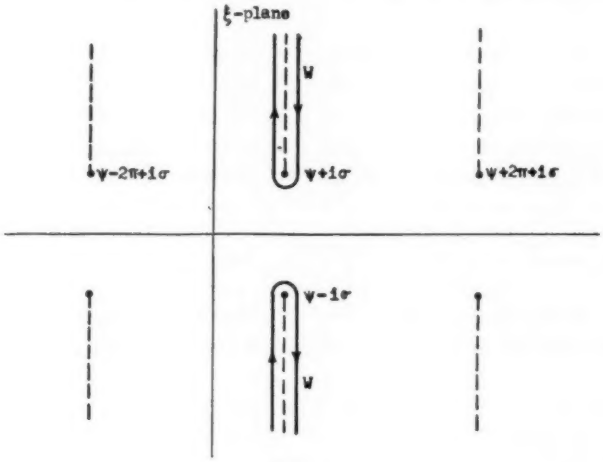

FIG. 4

For convenience, cut the ξ-plane by vertical lines proceeding from the branch points to $\pm i\infty$ as shown in Fig. 4. The functions $q(\xi)$, $t(\xi)$ are single-valued in this cut ξ-plane and take opposite signs on the opposite sides of each cut. We define $[s - \cos(\psi - \xi)]^{1/2}$ to be positive when ξ is real.

The basic multi-value dipole $Q(\psi_0)$ situated at a (real) point $\sigma = 0$, $\psi = \psi_0$ on the z-axis of our infinite sheeted Riemann space is defined as follows

$$Q(\psi_0) = \frac{1}{2\pi i}\int_W \frac{q(\xi)}{\xi - \psi_0}\, d\xi \tag{4.1}$$

where the path W of integration is the two-branched path indicated in Fig. 4. The function $Q(\psi_0)$ is a potential function in (σ, ψ) since differentiations can be performed under the integral sign, the path W being temporarily fixed.

For an n-sheeted Riemann space in which values of the angle ψ differing by $2n\pi$

would be identified, the corresponding n-valued dipole $Q(\psi_0)$ is

$$Q^{(n)}(\psi_0) = \frac{1}{2\pi i}\int_W \frac{q(\xi)}{2n} \cot \frac{\xi - \psi_0}{2n}\, d\xi. \tag{4.2}$$

Since $\cot (\xi-\psi_0)/2n$ is periodic in ξ of period $2n\pi$, and $q(\xi)$ is periodic of period 2π, it is easy to see that $Q^{(n)}(\psi_0)$ is periodic in ψ of period $2n\pi$.

We now briefly describe the properties of $Q(\psi_0)$, a similar description applying to $Q(\psi_0)$ (see [3], [4] for a complete analysis, the following being but a bare sketch). On account of the pole of the integrand in (4.1) at ψ_0 one sees that

$$Q(\psi_0) - q(\psi_0)$$

is regular as $\sigma\to 0$, $\psi\to\psi_0$ (where $\psi_0 \neq 2j\pi$). Thus $Q(\psi_0)$ behaves like a dipole near $\sigma=0$, $\psi=\psi_0$, but is regular near $\sigma=0$, $\psi=\psi_0+2k\pi$, $k=\pm 1, \pm 2, \cdots$ since the integrand is regular at these points.

In addition, there are still the poles of $q(\xi)$ itself at $\xi=2k\pi$, $k=0, \pm 1, \pm 2, \cdots$ due to the presence of the term $\cot \xi/2$ (see (3.1)). One easily sees that

$$Q(\psi_0) \to \frac{a}{2k\pi - \psi_0}$$

as $\sigma\to 0$ and $\psi\to 2k\pi$, $k=0, \pm 1, \pm 2, \cdots$ ($\psi_0 \neq 2j\pi$). Thus the function $Q(\psi_0)$ is not simply a multi-valued dipole at ψ_0 but it has certain non-vanishing values at ∞ ($\sigma=0$, $\psi=2k\pi$) in the various sheets and has no other singularities.

(A constant value at ∞ may be considered as a pole at ∞. If in the definition of $q(\xi)$ the term $\cot \xi/2$ were replaced by $\cot \psi_0/2$, the resulting function in (5.1) would be zero at the points at ∞; this would be a simple dipole. However, a simple dipole would lead to great difficulties, and our definition of $q(\xi)$ is in this case the correct generalization to the multi-sheeted case. This illustrates how careful one must be in the choice and analysis of the singularities, with special reference to the behavior at infinity.)

For the special case in which ψ_0 is a multiple of 2π, the integrand in (4.1) has a double pole. For example, $Q(0)$ has an integrand in (4.1) which has a double pole at $\xi=0$, and the residue there is

$$\frac{a}{2}\,\frac{\sin\psi}{s-\tau} + a\left[\frac{\partial}{\partial\xi}\left(\frac{s-\tau}{s-\cos(\psi-\xi)}\right)^{1/2}\right]_{\xi=0} = a\,\frac{\sin\psi}{s-\tau} = z.$$

Therefore

$$Q(0) - z \to 0 \quad \text{as} \quad \sigma\to 0,\ \psi\to 0 \tag{4.3}$$

and so $Q(0)$ behaves like z near the point at ∞ in the first sheet. A similar analysis applies to $Q(2k\pi)$.

In the n-sheeted case, it is of interest to consider the special case $n=1$ where we are in the original single-sheeted space. The integrand in (4.2) is then of period 2π, and we have

$$Q^{(1)}(\psi_0) = q(\psi_0) - a/2 \cot(\psi_0/2), \qquad \psi_0 \neq 0, \tag{4.4}$$

so $Q^{(1)}(\psi_0)$ differs from an ordinary dipole $q(\psi_0)$ by a constant. If in (4.4) we now let

$\psi_0 \to 0$, we obtain

$$Q^{(1)}(\psi_0) \to z \quad \text{as} \quad \psi_0 \to 0.$$

It may also be verified directly that

$$Q^{(1)}(0) = z. \tag{4.5}$$

5. Inversion of multi-valued dipoles. It is well known that in ordinary space the image of a dipole with respect to a sphere is another dipole situated at the inverse point. In this section the theorem will be extended to multi-sheeted spaces.

THEOREM 5.1. *The function* $Q(\psi_0) - Q(\psi_0')$ *has vanishing normal derivative on the spherical cap* $\psi = \text{constant} = c$ *if* $\psi_0 + \psi_0' = 2c$.

Proof. The stream function $T(\psi_0)$ corresponding to $Q(\psi_0)$ is

$$T(\psi_0) = \frac{1}{2\pi i} \int_W \frac{t(\xi)}{\xi - \psi_0}\, d\xi \tag{5.1}$$

where $t(\xi)$ is the stream function corresponding to $q(\xi)$, obtained from (3.2) by replacing ψ_0 by ξ. It is required to show merely that $T(\psi_0) - T(\psi_0')$ is constant on $\psi = c$.

Make the substitution $\xi - \psi = \zeta$. Then

$$t(\xi) = \frac{a^2}{2} \frac{\sinh^2 \sigma}{(s - \tau)^{1/2}(s - \cos \zeta)^{3/2}} \tag{5.2}$$

$$T(\psi_0) = \frac{1}{2\pi i} \int_{W'} \frac{t(\xi)}{\zeta + \psi - \psi_0}\, d\zeta \tag{5.3}$$

where W' is the path in the ζ-plane shown in Fig. 5. By the symmetry of the path W', the lower path W_l' is obtained from the upper path W_u' by replacing ζ by $-\zeta$. The

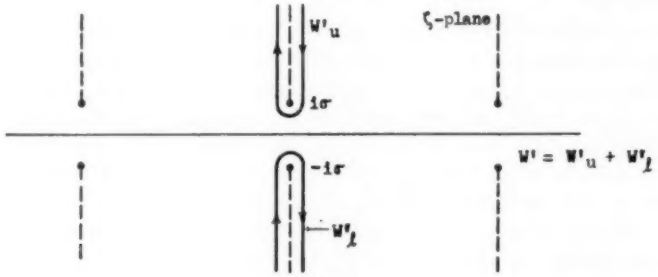

FIG. 5

function $t(\xi)$ is by (5.2) an even function of ζ, and (5.3) can be written in the form

$$T(\psi_0) = \frac{1}{2\pi i} \int_{W'_u} t(\xi) \left(\frac{1}{\zeta + (\psi - \psi_0)} + \frac{1}{\zeta - (\psi - \psi_0)} \right) d\zeta. \tag{5.4}$$

Likewise for $T(\psi_0')$. Evaluating at $\psi = c$, and noting that the hypothesis $\psi_0 + \psi_0' = 2c$ can be written $c - \psi_0 = -(c - \psi_0')$, eq. (5.4) shows immediately that $T(\psi_0) - T(\psi_0') = 0$ when $\psi = c$, and the theorem is proved.

THEOREM 5.2. *The function* $Q(\psi_0)+Q(-\psi_0)$ *vanishes on* $\psi=0$. *In particular* $Q(0)$ *vanishes on* $\psi=0$.

Proof. For $\psi=0$ the function $q(\xi)$ is an odd function of ξ by (3.1). The same argument as above yields $Q(\psi_0)=-Q(-\psi_0)$ for $\psi=0$.

6. Construction of the potential function for a symmetrical lens. We are now prepared to apply the method of images described in Sec. 2. For the sake of convenience we shall first consider the case of a symmetrical lens, the two spherical portions having the equation $\psi=\beta$ and $\psi=-\beta$ where the exterior angle between the two intersecting spheres is 2β. The radius of the circle of intersection of the spheres is a, the semi-thickness of the lens is B, and the radius of the two symmetrical spheres is A, as indicated in Fig. 6. The motion of the lens is in the negative z-direction.

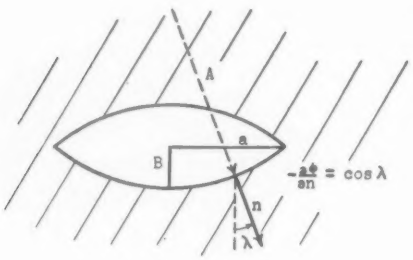

FIG. 6

Let Φ be the desired potential function, and set

$$V=\Phi-z. \tag{6.1}$$

Then V is a potential satisfying the following conditions

$$-\frac{\partial V}{\partial n}=0 \quad \text{on} \quad \psi=\beta \tag{6.2}$$

$$-\frac{\partial V}{\partial n}=0 \quad \text{on} \quad \psi=-\beta \tag{6.3}$$

and

$$V \text{ behaves like } -z \text{ at } \infty. \tag{6.4}$$

The potential function V corresponds to the flow about a fixed lens, the fluid having unit velocity at ∞.

We shall construct V by the method of images, using the infinite sheeted Riemann space described by $0\leqq\sigma\leqq\infty$, $-\infty<\psi<\infty$. (It may be pointed out that if the angle $\beta=n\pi/2q$ where n, q are integers, it would be sufficient to use an n-sheeted Riemann space in place of the infinite sheeted one.) First start with $-Q(0)$ which has the correct behavior (6.4) as $\sigma\to0$, $\psi\to0$ (by (4.3)). To satisfy (6.2), one must add $Q(2\beta)$ by theorem (5.1); and to satisfy (6.3) one must add $Q(-2\beta)$. But $Q(-2\beta)$ destroys the correct boundary condition on $\psi=\beta$, and likewise for $Q(2\beta)$ on the spherical cap $\psi=-\beta$. To remedy this one must add $-Q(4\beta)$ by Theorem 5.1, and likewise $-Q(-4\beta)$. Continuing this process indefinitely, one obtains

$$V=\sum_{k=-\infty}^{\infty}(-1)^{k+1}Q(2k\beta). \tag{6.5}$$

The convergence of this sum follows directly from its evaluation below.

We may write (6.5) in the forms

$$V=\sum_{k=0}^{\infty}(-1)^{k+1}[Q(-2k\beta)-Q(2(k+1)\beta)] \tag{6.6}$$

$$V = \sum_{k=0}^{\infty} (-1)^{k+1}[Q(2k\beta) - Q(-2(k+1)\beta)]. \tag{6.7}$$

By virtue of Theorem 5.1, eqs. (6.6) and (6.7) show that (6.2) and (6.3) are satisfied respectively. From the appearance of $-Q(0)$ in (6.5) we see that (6.4) is also satisfied.

Substituting from (4.1) we obtain

$$V = \frac{1}{2\pi i}\int_W q(\xi) \sum_{k=-\infty}^{\infty} (-1)^{k+1} \frac{1}{\xi - 2k\beta} d\xi$$

or

$$V = -\frac{1}{2\pi i}\int_W q(\xi) \frac{\pi}{2\beta} \csc \frac{\pi\xi}{2\beta} d\xi. \tag{6.8}$$

For the desired potential function Φ, we have from (6.1) and (4.5),

$$\Phi = -\frac{1}{2\pi i}\int_W q(\xi)\left(\frac{\pi}{2\beta} \csc \frac{\pi\xi}{2\beta} - \frac{1}{2}\cot\frac{\xi}{2}\right) d\xi. \tag{6.9}$$

Formulas (6.8) or (6.9) are the final expressions for V or Φ respectively.

The integrals may be expressed in real form by integrating by parts and then moving the path W into the cuts which it surrounds. The results are

$$V = \frac{A \sin\beta}{2\beta}\int_\sigma^\infty \sqrt{\frac{s-\tau}{\cosh u - s}}\, \mathrm{Im}\left[\cot\frac{\xi}{2} \csc\frac{\pi\xi}{2\beta} + \frac{\pi}{\beta}\csc\frac{\pi\xi}{2\beta}\cot\frac{\pi\xi}{2\beta}\right] du,$$

where $\xi = \psi + iu$, and an analogous expression for Φ.

7. Virtual mass of the fluid. The kinetic energy of the fluid is $(M/2) \times$ (velocity of object)^2, where

$$M = \rho\iiint (-\operatorname{grad}\Phi)^2 dx\,dy\,dz = -\iint_{\text{lens}} \Phi \frac{\partial\Phi}{\partial n} dS,$$

and ρ is the density of the fluid, n is a normal directed into the fluid. The contribution due to a surface integral over a surface which approaches ∞ vanishes. Using the boundary condition for $-\partial\Phi/\partial n$ on the surface of the lens,

$$M = \rho\iint_{\text{lens}} \Phi \cos\lambda\, dS, \tag{7.1}$$

where λ is the angle between the normal and the direction of motion, indicated in Fig. 6. This quantity M is called the "virtual mass" or the "added mass" of the fluid. The same formula (7.1) is obtained if the virtual mass is based on the vertically downward momentum of the fluid.

The virtual mass M is more easily calculated by obtaining a different form for (7.1).* Formula (7.1) can be written (motion is in the negative z-direction)

$$M = -\rho\iint_{\text{lens}} \Phi \frac{\partial z}{\partial n} dS,$$

* The longer calculation required in (7.1) was carried out in the original OSRD report with the same result.

since $-\partial z/\partial n = \cos\lambda$. Green's formula,

$$\iint \left(u \frac{\partial v}{\partial n} - v \frac{\partial u}{\partial n} \right) dS = 0,$$

can be applied to the two potential functions Φ, z yielding

$$M = -\rho \iint_{\text{lens}} z \frac{\partial \Phi}{\partial n} \, dS + \rho \iint_{\infty} \left(\Phi \frac{\partial z}{\partial n} - z \frac{\partial \Phi}{\partial n} \right) dS \tag{7.2}$$

where the last integral is taken over a surface which approaches ∞, for example a sphere of radius R where $R \to \infty$. At ∞, Φ behaves like a dipole and we may set

$$\Phi \sim D \cdot \cos\theta / R^2 \quad \text{at} \quad \infty \tag{7.3}$$

where $(\cos\theta)/R^2$ is a unit dipole at th origin pointing in the direction of motion of the lens, and D is the strength of the dipole equivalent ta ∞ to Φ. Then on a sphere of radius R, $z = -R\cos\theta$, and a simple calculation gives $4\pi\rho D$ as the value of the second integral in (7.2).

For the first integral in (7.2) over the surface of the object, we have

$$-\frac{\partial \Phi}{\partial n} = \cos\lambda$$

and

$$-\rho \iint_{\text{lens}} z \frac{\partial \Phi}{\partial n} \, dS = \rho \iint z \cos\lambda \, dS = -\rho \cdot \text{volume of lens}$$

since $\cos\lambda \, dS = \pm dx\,dy$, the sign depending on the sign of $\cos\lambda$. Substitution of both evaluations into (7.2) yields the final formula, valid for any symmetrical object moving in an infinite fluid:*

$$M = 4\pi\rho D - \text{mass of fluid displaced by object} \tag{7.4}$$

where D is the "dipole strength" at ∞ of the potential function Φ as defined in (7.3).

It remains to evaluate the dipole strength D for the particular lens flow. The dipole $q(\xi)$ from its very construction has strength

$$a^3 \left| \csc^3(\xi/2) \right| /2 \tag{7.5}$$

where $\left|\csc^3(\xi/2)\right|$ is the piecewise analytic function of the complex variable ξ equal to $\pm\csc^3(\xi/2)$, the sign being determined in such a way that $\left|\csc^3(\xi/2)\right|$ is positive for real ξ (there are vertical cuts extending from $2n\pi - i\infty$ to $2n\pi + i\infty$). This may also be verified directly from (3.1) or (3.2).

The dipole strength of Φ can now be obtained from (6.9). The integrand has no singularity near $\xi = 0$ or near $\xi = \psi$ (we are considering $\psi \to 0$), so that the path W may be deformed into the path W'' indicated in Fig. 7a. Inserting the dipole strength (7.5) of $q(\xi)$, one obtains for the dipole strength D of Φ at ∞:

$$D = -\frac{1}{2\pi i} \int_{W''} \frac{a^3}{2} \left| \csc^3(\xi/2) \right| \left(\frac{\pi}{2\beta} \csc\frac{\pi\xi}{2\beta} - \frac{1}{2} \cot\frac{\xi}{2} \right) d\xi \tag{7.6}$$

* The same formula applies for any (asymmetrical) object but then D is the component of the dipole strength at ∞ in the direction of motion.

where now the diagram in the ξ-plane is indicated by Fig. 7b. The second term in the integrand involving $\cot \xi/2$ can be integrated explicitly and has the value zero. Further, the path W'' reduces to W_1'' by replacing ξ by $-\xi$. Making the change of variable $\pi\xi/2\beta = z$, we have

$$D = \frac{a^2}{2\pi i}\int_L \csc^3 \frac{\beta z}{\pi} \csc z dz \tag{7.7}$$

where the path W_1'' is henceforth denoted by L and is a vertical line slightly to the left of the imaginary axis.

The virtual mass can now be obtained from (7.4), (7.7) and

$$\text{mass of displaced fluid} = \tfrac{8}{3}\pi\rho A^3(2 - \cos\beta)\cos^4\beta/2. \tag{7.8}$$

The integral in (7.7) is evaluated for angles β which are rational multiples of π in Sec. 9. The results for the virtual mass as a function of the thickness of the lens are exhibited in dimensionless form in Figs. 8 and 9. In Fig. 8 we have plotted $M/\pi\rho A^3$ as a function of B/A, where A is the radius of the spheres bounding the lens and B is the half thickness of the lens. The case of completely separated spheres, $B/A > 2$, is included in the graph (see App. C). It is also of interest to emphasize the presentation surface of the lens, and for this purpose we have plotted $M/\pi\rho a^3$ as a function of B/a in Fig. 9, where a is the radius of the sharp circular rim of the lens. For $B/a > 1$, the lens has a figure eight cross section and the presentation radius is preferably A instead of a, so that Fig. 8 should be used for this range.

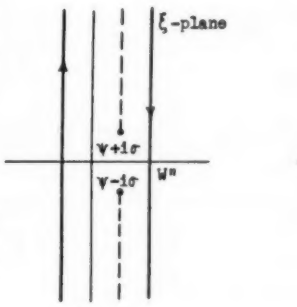

FIG. 7a

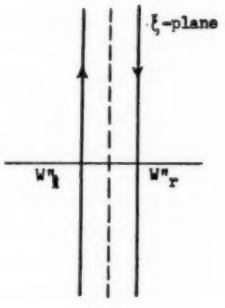

FIG. 7b

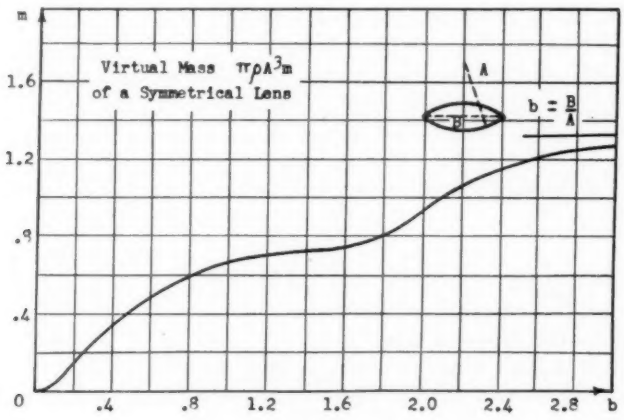

FIG. 8

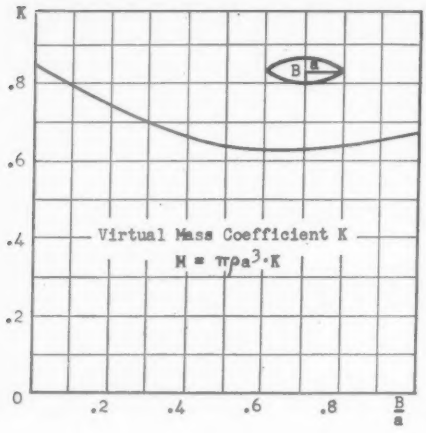

FIG. 9

8. Potential function and virtual mass for an arbitrary lens. We construct the velocity potential of the flow around an asymmetrical lens bounded by the spherical caps $\psi=\alpha-\beta$ and $\psi=-\beta$, where β and $\alpha-\beta$ are both positive. We suppose that the lens is immersed in an infinite fluid and that it is moving downward (i.e. in the negative z-direction) with unit velocity. The symmetrical lens is obtained by taking $\alpha=2\beta$, and this is the case which has been discussed in the main body of the paper.

The potential function describing the flow satisfies conditions (6.1)–(6.4) as in Sec 6, but. in place of (6.2), we have

$$-\frac{\partial V}{\partial n}=0 \quad \text{on} \quad \psi=\alpha-\beta \quad \text{and on} \quad \psi=-\beta. \tag{8.1}$$

In order to illustrate a remark made in Sec. 6, namely that the n-fold Riemann space may be used if the intersection angle, here equal to α, is rational, we suppose that $\alpha=n\pi/m$. The formulas which we obtain are then valid for all α by continuity.

The solution is the potential function

$$\begin{aligned} V &= \sum_{k=1}^{m}\left[Q^{(n)}(2k\alpha-2\beta)-Q^{(n)}(-2k\alpha)\right] \\ &= \sum_{k=1}^{m}\left[Q^{(n)}(2(k+1)\alpha-2\beta)-Q^{(n)}(-2k\alpha)\right] \end{aligned} \tag{8.2}$$

making use of the fact that $Q^{(n)}(\psi_0)$ is periodic in ψ_0 of period $2n\pi$. The first form of (8.2) shows by Theorem 5.1 that V has vanishing normal derivative on $\psi=-\beta$, while the second form yields the vanishing of the normal derivative on $\psi=\alpha-\beta$. Furthermore, at ∞ in the first sheet ($\sigma=0$, $\psi=0$), all the terms in the summation are regular except $-Q^{(n)}(-2m\alpha)=-Q^{(n)}(-2n\pi)=-Q^{(n)}(0)$ which behaves at ∞ like $-z+0(1)$ (compare (4.3)). Substituting from (4.2) into (8.2) and evaluating the sums, we obtain the formula

$$V=\frac{1}{2\pi i}\int_{W} q(\xi)\,\frac{\pi}{2\alpha}\left(\cot\frac{\pi(\xi+2\beta)}{2\alpha}-\cot\frac{\pi}{2\alpha}\right)d\xi. \tag{8.3}$$

The desired potential function $\Phi=V+z$.

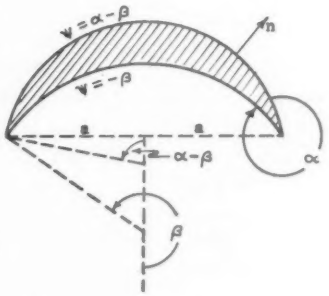

FIG. 10

Formula (8.3) reduces to the old formula (6.8) in the special case when $\alpha=2\beta$. It is interesting to note that (8.3) is built out of two terms of the type

$$\overline{Q}(\psi_0) = \frac{1}{2\pi i}\int_W q(\xi)\,\frac{\pi}{2\alpha}\cot\frac{\pi(\xi-\psi_0)}{2\alpha}\,d\xi. \tag{8.4}$$

According to Theorem 5.1, the image of $-\overline{Q}(0)$ relative to $\psi=-\beta$ is $\overline{Q}(-2\beta)$, while its image relative to $\psi=\alpha-\beta$ is $\overline{Q}(2\alpha-2\beta)$. But $\overline{Q}(-2\beta)=\overline{Q}(2\alpha-2\beta)$ since $\overline{Q}(\psi_0)$ is periodic in ψ_0 of period 2α. Thus the potential function

$$V = -\overline{Q}(0) + \overline{Q}(-2\beta) \tag{8.5}$$

satisfies the boundary conditions on both $\psi=-\beta$ and $\psi=\alpha-\beta$, and has the required behavior at ∞. It is therefore the solution to our problem. This is exactly (8.3).

The expression $\overline{Q}(\psi_0)$ in (8.4) might be interpreted as a fundamental dipole in a space obtained by identifying angles ψ which differ by 2α. This would be analogous to the two-dimensional case in which V could be extended across $\psi=-\beta$ into the region between $\psi=-\beta$ and $\psi=-\alpha-\beta$ by inversion and then noting that this extension is likewise the extension across $\psi=\alpha-\beta$ if one identifies $\psi=\alpha-\beta$ with $\psi=-\alpha-\beta$. But the simple extension by inversion does not apply in the three-dimensional case but rather a more complicated continuation. In fact $\overline{Q}(\psi_0)$ in (8.4) is not single valued in the space obtained by identifying angles ψ which differ by 2α. Thus the analogy with the two-dimensional case given by formulas (8.4), (8.5) is not exact, but is nevertheless noteworthy.

The calculation of the virtual mass is similar to that of Sec. 7. The result is

$$M = M_1 - M_2 \tag{8.6}$$

where

$$M_1 = 2\pi\rho a^3 \sin\frac{\pi\beta}{\alpha}\,\frac{1}{2\pi i}\int_L \csc^3\frac{\alpha z}{\pi}\csc z\left\{\csc\left(z+\frac{\pi\beta}{\alpha}\right)-\csc\left(z-\frac{\pi\beta}{\alpha}\right)\right\}dz \tag{8.7}$$

and where

$$M_2 = \frac{1}{6}\pi\rho a^3\Bigg[(2-\cos\beta)\cot\frac{\beta}{2}\csc^2\frac{\beta}{2} + (2-\cos(\alpha-\beta))\cot\frac{\alpha-\beta}{2}\csc^2\frac{\alpha-\beta}{2}\Bigg] \tag{8.8}$$

is the mass of the fluid displaced by the lens.

9. The symmetrical lens. Virtual mass in special cases. (i) For the determination of the virtual mass there remains but to evaluate the integral in (7.7), and this can be done in terms of residues when the angle β is a rational multiple of π. Suppose that

$$\beta = \frac{n\pi}{q} \tag{9.1}$$

where n and q are relatively prime integers and $n<q$. The integral to be evaluated is then

$$I = \frac{1}{2\pi i}\int_L \csc^3\frac{nz}{q}\csc z\,dz \tag{9.2}$$

where we repeat that L is a vertical line slightly to the left of the imaginary axis.

Let δ be a fixed positive number and introduce the factor $e^{-\delta z}$ in the integrand in (9.2). Then

$$\frac{1}{2\pi i}\int_L \csc^3\frac{nz}{q}\csc z e^{-\delta z}dz = -\text{ sum of residues to the right of } L$$
$$= -\frac{1}{1-(-1)^{n+q}e^{-\delta q\pi}} \times \begin{matrix}\text{sum of residues}\\ \text{between 0 and } q\pi.\end{matrix} \tag{9.3}$$

Calculating the residues between 0 and $q\pi$, letting $\delta\to 0$ and observing that the coefficient of $1/\delta$ must vanish, we obtain two different results according as $n+q$ is odd or even. If $n+q$ is odd,

$$I = \frac{1}{2}\Big[\sum_{j=1}^{q-1}(-1)^{j-1}\csc^3\frac{jn\pi}{q} + \frac{q}{2n}\sum_{k=1}^{n-1}(-1)^{k-1}\left\{\csc z + \frac{q^2}{n^2}\frac{d^2}{dz^2}(\csc z)\right\}_{z=kq\pi/n}\Big], \tag{9.4}$$

and if $n+q$ is even,

$$I = \frac{1}{q\pi}\Big[\left(\frac{q^3}{6n^3}+\frac{q}{2n}\right) + \pi\sum_{j=1}^{q-1}(-1)^{ij}\csc^3\frac{jn\pi}{q} + \frac{q}{2n}\sum_{k=1}^{n-1}(-1)^{k}\left\{z\csc z + \frac{q^2}{n^2}\frac{d^2}{dz^2}(z\csc z)\right\}_{z=kq\pi/n}\Big]. \tag{9.5}$$

The computations based on these formulas were used in plotting the graph of Fig. 8.

It is interesting to note that formulas (9.4), (9.5) involve angles which are multiples of $q\pi/n$ as well as those which are multiples of $n\pi/q$. These trigonometric sums, which would normally depend in a highly discontinuous manner on the angle $\beta = n\pi/q$, do indeed depend continuously since they are special cases of the integral (9.2).

For the purpose of checking, it is worth while to calculate the virtual mass in certain easy special cases where it can be calculated by other means.

(ii) If $\beta=\pi/2$, eq. (9.4) for $n=1$, $q=2$ gives $I=1/2$, or from (7.4), (7.7), (7.8)

$$M = \tfrac{2}{3}\pi\rho A^3.$$

This trivial case corresponds to the flow about a sphere in an infinite fluid, and the virtual mass is then known to be as stated.

(iii) If we let $\beta\to 0$, we obtain from (7.7) and the fact that $a = A\sin\beta$,

$$D \to \frac{\pi^3 A^3}{2\pi i}\int_L \frac{\csc z}{z^3}dz = A^3\sum_{j=1}^{\infty}(-1)^{j-1}/j^3.$$

Substituting in (7.4), (7.7), (7.8), the virtual mass of the fluid, when the two spheres are tangent ($\beta=0$), is

$$M = 4\pi\rho A^3\Big[\sum_{j=1}^{\infty}(-1)^{j-1}/j^3 - 2/3\Big]. \tag{9.6}$$

This value agrees with the limit obtained when completely separated spheres ap-

proach tangency (see App. C) and so is a consequence of known formulas (see [2], p. 447).

(iv) Taking $\beta=\pi$, the virtual mass M is obviously zero. However if we keep a (instead of A) as fixed, then we obtain the virtual mass for a circular disc of radius a. Taking $\beta=\pi$ in (7.4), (7.7), (7.8), we obtain

$$M = \frac{4\pi\rho a^3}{2\pi i}\int_L \csc^4 z\,dz = \tfrac{8}{3}\rho a^3.$$

This is also a known result (see [2], p. 428).

10. The symmetrical lens. Derivative of the virtual mass. For the application of the lens flow to the impact of a sphere on a fluid surface, it is necessary to calculate

$$\frac{dm}{db} \quad \text{where} \quad m = M/\pi\rho A^3, \quad b = B/A.$$

This can be done by differentiating formulas (7.4), (7.7), (7.8), leading to another trigonometric contour integral. This integral can then be evaluated for $\beta=n\pi/q$ as in the preceding section, with two cases appearing, according as $n+q$ is odd or even. We shall not write down these formulas since they are complicated and are similar to those in the preceding section. Just the remark should be made that in equating the coefficient of $1/\delta$ to zero one obtains evaluations of certain trigonometric sums, involving angles which are multiples of $q\pi/n$ and $n\pi/q$, which appear at first glance quite startling.

On the basis of these formulas, a graph of dm/db was prepared and is shown in Fig. 11.

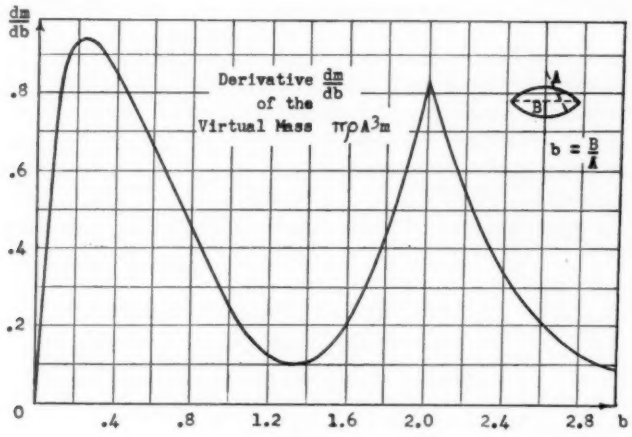

FIG. 11

APPENDIX A

Figure eight lens with exterior angle $\alpha=\pi/m$. The elementary process of successive images of dipoles fails in general without the introduction of multi-sheeted spaces. But this method succeeds for the special case of a lens with exterior angle $\alpha=\pi/m$,

in which case the procedure described in Sec. 2 terminates after m steps. On the mth step, the same dipole corrects the boundary conditions on both spheres K and K' simultaneously, and the total number of dipoles is $2m-1$. An example is a figure eight lens with $\alpha=\pi/2$, which has three dipoles situated at the points O, O', Q indicated in Fig. 12. (By inversion relative to K or to K', the circle $OR\,O'S$ inverts into the straight line RQS. Thus Q is the inverse of O' relative to K and also of O relative to K'.) The dipoles are directed vertically downward and their strengths at O, O' and Q are

$$A^3/2, \quad A'^3/2 \quad \text{and} \quad -A^3A'^3/2(\overline{OO'})^3 = -a^3/2 \text{ respectively.}$$

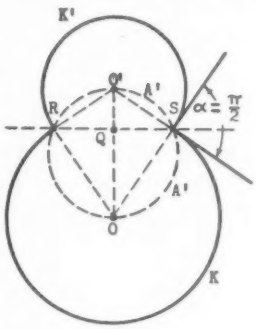

FIG. 12

For the general case in which $\alpha=\pi/m$, it is convenient to use the angular coordinate ψ. Let the equations of the two spherical caps K and K' be $\psi=-\alpha/2+\gamma$ and $\psi=\alpha/2+\gamma$ respectively, where $-\alpha/2<\gamma<\alpha/2$. The dipoles are situated on the axis of symmetry at the points

$$P_\nu; \quad \psi_\nu = -(\nu+1)\alpha + [1+(-1)^\nu]\gamma, \qquad \nu = 0, \cdots, m-1$$
$$P'_\nu; \quad \psi'_\nu = \quad (\nu+1)\alpha + [1+(-1)^\nu]\gamma, \qquad \nu = 0, \cdots, m-1.$$

The points P_{m-1} and P'_{m-1} are identical, since $\psi'_{m-1}-\psi_{m-1}=2\pi$. The strengths of the dipoles at P_ν and P'_ν are $(-1)^{\nu-1}\overline{SP}_\nu^3/2$, where S is a point on the sharp circular rim of intersection of the two spherical caps and $\overline{SP}_\nu$ is the distance from P_ν to S.

The expression for the velocity potential is now easily written down in terms of these $(2m-1)$ dipoles, and the virtual mass obtained from formula (7.4).

APPENDIX B

Spherical shell and figure eight lens with exterior angle $\alpha=2\pi/m$. In the special case when the exterior angle $\alpha=2\pi/m$, the velocity potential V (or Φ) can be expressed in terms of elementary functions. According to Sec. 8 all that is required is a two-sheeted Riemann space and the fundamental dipoles can then be explicitly integrated.

It is much simpler to obtain an expression for the Stokes' stream function instead of the velocity potential. Let $T^{(n)}(\psi_0)$, S, χ be the Stokes' stream functions conjugate to $Q^{(n)}(\psi_0)$, V, Φ respectively. Corresponding to (4.2), (8.3) and $\Phi=V+z$ we have

$$T^{(n)}(\psi_0) = \frac{1}{2\pi i}\int_W \frac{t(\xi)}{2n}\cot\frac{\xi-\psi_0}{2n}\,d\xi \tag{B.1}$$

$$S = \frac{1}{2\pi i}\int_W \frac{\pi t(\xi)}{2\alpha}\left[\cot\frac{\pi(\xi+2\beta)}{2\alpha} - \cot\frac{\pi\xi}{2\alpha}\right]d\xi \tag{B.2}$$

$$\chi = S + r^2/2 \tag{B.3}$$

where $t(\xi)$ is obtained from $t(\psi_0)$ in (3.2) by replacing ψ_0 by ξ, and $r^2/2$ is the stream function conjugate to z. In any numerical work these have a much simpler form than the corresponding formulas for the potential.

In the integrand the substitution $\xi=\psi+\zeta$ can be made, and the lower half W'_l of the path of integration can be reduced to the upper half W'_u by the substitution $\zeta\to-\zeta$. There results

$$T^{(n)}(\psi_0) = \frac{a^2 \sinh^2 \sigma}{4\pi n i (s-\tau)^{1/2}} \int_{W'_u} \frac{1}{(s-\cos\zeta)^{3/2}} \cdot \frac{\sin(\zeta/n)d\zeta}{\cos[(\psi-\psi_0)/n] - \cos(\zeta/n)}. \tag{B.4}$$

Consider now the case $n=2$. Make the substitution $\cos(\zeta/2)=\cosh(\sigma/2)\cosh\eta$ and evaluate the integral. The result is

$$T^{(2)}(\psi_0) = t(\psi_0)\frac{2}{\pi}\arctan\left[\left\{\frac{\cosh(\sigma/2)+\cos[(\psi-\psi_0)/2]}{\cosh(\sigma/2)-\cos[(\psi-\psi_0)/2]}\right\}^{1/2}\right] + \frac{2^{1/2}a^2\sinh^2(\sigma/2)\cos[(\psi-\psi_0)/2]}{\pi(s-\tau)^{1/2}[s-\cos(\psi-\psi_0)]} \tag{B.5}$$

where $t(\psi_0)$ is the fundamental dipole in (3.2).

The arc tangent factor in (B.5) illustrates the difference in behavior of $T^{(2)}(\psi_0)$ between $\psi\to\psi_0$, $\sigma\to 0$ and $\psi\to\psi_0+2\pi$, $\sigma\to 0$. In the former situation, the factor approaches 1 while in the latter it approaches 0. The second term in (B.5) just cancels the remaining singularity, with the result that $T^{(2)}(\psi_0)-t(\psi_0)$ is regular as $\psi\to\psi_0$, $\sigma\to 0$ while $T^{(2)}(\psi_0)$ itself is regular as $\psi\to\psi_0+2\pi$, $\sigma\to 0$.

We may now obtain the stream function for the flow about a figure eight lens with $\alpha=2\pi/m$. For the stream function S conjugate to V we have a formula analogous to (8.2):

$$S = \sum_{k=1}^{m}\left[T^{(2)}(4k\pi/m - 2\beta) - T^{(2)}(-4k\pi/m)\right]. \tag{B.6}$$

The case $m=1$ is especially noteworthy since this corresponds to the flow about spherical shell (exterior angle 2π):

$$S = T^{(2)}(-2\beta) - T^{(2)}(0), \qquad \alpha = 2\pi \text{ (spherical shell)}. \tag{B.7}$$

For the stream function χ, conjugate to Φ, we have

$$\chi = T^{(2)}(-2\beta) + T^{(2)}(2\pi), \qquad \alpha = 2\pi \text{ (spherical shell)} \tag{B.8}$$

since $r^2/2 = T^{(1)}(0) = T^{(2)}(0)+T^{(2)}(2\pi)$. The flow about a spherical shell has also been obtained by Basset [5] using spherical coordinates.

For the sake of completeness, we include here the formula for $Q^{(2)}(\psi_0)$. It is

$$Q^{(2)}(\psi_0) = q(\psi_0)f(\psi_0) + \frac{a}{\pi\sqrt{2}}\frac{(s-\tau)^{1/2}\sin[(\psi-\psi_0)/2]}{s-\cos(\psi-\psi_0)} - \frac{a}{2}\csc(\psi_0/2)f(0) + \frac{a}{4}\tan(\psi_0/4),$$

where $f(\psi_0)$ is the arc tangent factor of $t(\psi_0)$ in (B.5),

$$f(\psi_0) = \frac{2}{\pi}\arctan[\quad].$$

The potential function arising from a pole (instead of a dipole) of suitable strength has the form

$$\frac{(s-\tau)^{1/2}}{[s-\cos(\psi-\psi_0)]^{1/2}} f(\psi_0).$$

In particular if the pole is at infinity ($\psi_0=0$), its form is

$$f(0).$$

This last expression is a potential function, equal to 1 at infinity in the first sheet ($\psi=0$), equal to 0 at infinity in the second sheet ($\psi=2\pi$) and having no singularities in the finite.

Appendix C

Completely separated spheres. Although the case of completely separated spheres can be treated by the method of images described in Section 2, and the virtual mass computed from (7.4), nevertheless it is of interest to trace a closer relationship to the case of intersecting spheres. It will be shown that the integral formulas for lens flow carry over to separated spheres by merely allowing the coordinates σ, ψ and the geometrical quantities a, β, α to be pure imaginary.

For simplicity we shall discuss only the symmetrical case, the extension to the asymmetrical case being clear. Let the two spheres be of radius A, as indicated in Fig. 13, moving at unit velocity in the direction of their line of centers. The distance between the centers of the spheres is $2(B-A)$. As in the lens case, set

$$B = A(1+\cos\beta), \qquad a = A\sin\beta \tag{C.1}$$

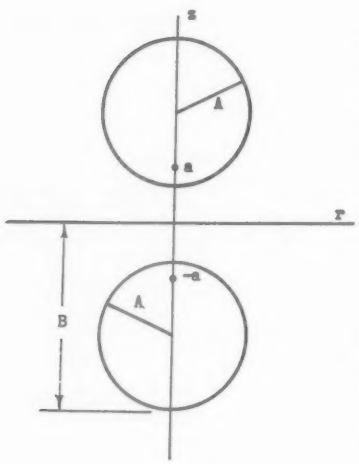

FIG. 13

where now β and a are pure imaginary. The points marked a and $-a$ in the diagram are mutually inverse with respect to both spheres. As in the lens case, the dipoles describing the flow are situated at the points $r=0, z=a\cot k\beta$ for $k=\pm1, \pm2, \cdots$ with strengths $(-1)^{k+1}\left|a^3\csc^3(k\beta)/2\right|$ respectively.*

The virtual mass M can be obtained from formula (8.4) with the result

$$m = \frac{M}{\pi\rho A^3} = 4\sum_{k=1}^{\infty}(-1)^{k+1}\frac{\sin^3\beta}{\sin^3 k\beta} - \frac{8}{3}, \tag{C.2}$$

where it should be recalled that β is imaginary and so the series is rapidly convergent. In particular, if the spheres approach tangency, $\beta=0$ and the result (9.6) is obtained.

As β varies from π to 0 along the real axis, the spheres first form a thin lens which thickens until the spheres are tangent. Then, as β varies from 0 to $i\infty$ along the imaginary axis, the spheres separate and move farther and farther

* We remark that in any problem involving successive inversions relative to two equal spheres, the introduction of β by means of (C.1) is a simplifying device. A similar device applies to any two spheres, and will simplify such formulas as in [1], pp. 131–133 and [2], p. 420. This remark also applies to the electrostatic analogue, as in [3], pp. 642–649.

apart. The virtual mass m and its derivative dm/dq vary continuously, but we see from Fig. 11 that d^2m/db^2 has a discontinuity at $\beta=0$ ($b=1+\cos\beta=2$).

We may also obtain the potential function in a form analogous to (6.8) or (6.9). Allow the coordinates σ, ψ as well as a, β to be pure imaginary. (More generally, complex values of σ, ψ, a, β may also be selected.) Let r, z be connected with σ, ψ by the same relations,

$$r = \frac{a \sinh \sigma}{\cosh \sigma - \cos \psi}, \qquad z = \frac{a \sin \psi}{\cosh \sigma - \cos \psi}.$$

The coordinates r, z are still real when σ, ψ, a are all pure imaginary. In order to have $r \geqq 0$, we suppose that $0 \leqq \operatorname{Im} \sigma \leqq \pi$ while $\operatorname{Im} \psi$ varies from $-\infty$ to $+\infty$. The locus $\psi = -\beta$ is the sphere of radius A and center at $r=0$, $z=-A\cos\beta$. The coordinates $\operatorname{Im}\sigma$, $\operatorname{Im}\psi$ and the angle θ of rotation form the so-called bipolar coordinate system in space.

The function defined by (3.1) remains a potential function even though σ, ψ are pure imaginary. The branch points of $q(\xi)$ as a function of the complex variable ξ are still located at $\psi - i\sigma + 2k\pi$, $\psi + i\sigma + 2k\pi$ but all of these now lie on the line $\operatorname{Im}\xi = \operatorname{Im}\psi$. The ξ-plane can be cut along the straight line segments connecting the two adjacent branch points $\psi - i\sigma + 2k\pi$ and $\psi + i\sigma + 2(k+1)\pi$ (see Fig. 14). Select the path $\tilde{W}$ as indicated in Fig. 14; $\tilde{W}$ can also be considered as surrounding all the individual cuts as well.

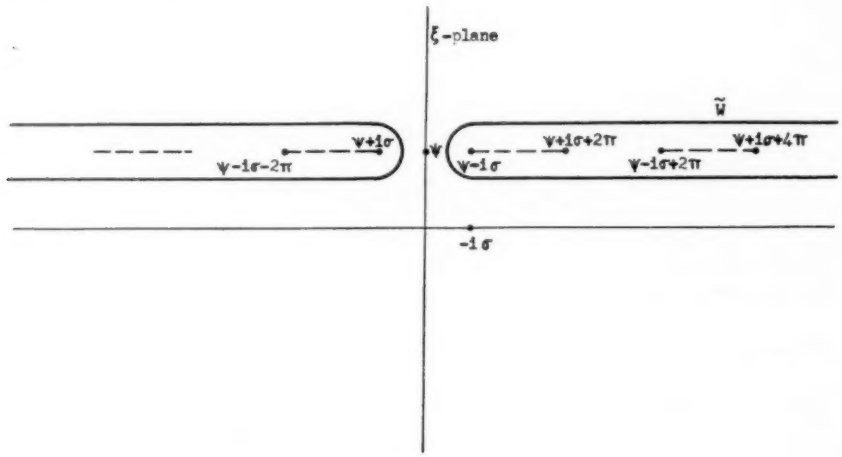

FIG. 14

The potential function V (or Φ) can now be constructed as in the lens case, with the same form as in (6.8):

$$V = -\frac{1}{2\pi i}\int_{\tilde{W}} q(\xi)\,\frac{\pi}{2\beta}\csc\frac{\pi\xi}{2\beta}\,d\xi. \tag{C.3}$$

The formulas for the virtual mass M is the same as (7.4), (7.7) except that the mass

of the fluid displaced is merely $\frac{8}{3}\pi\rho A^3$. The integral in (7.7) can then be evaluated by residues for β pure imaginary, with the result already obtained in (C.2).

REFERENCES

[1] H. LAMB, *Hydrodynamics*, 6th ed., MacMillan, 1932.

[2] L. M. Milne-Thomson, *Theoretical hydrodynamics*, MacMillan, 1939.

[3] P. Frank and R. v. Mises, *Die Differential- und Integralgleichungen der Mechanik und Physik*, Rosenberg Publ., N. Y. C., 1943, vol. 2, pp. 808–875, by A. Sommerfeld, "Theorie der Beugung."

[4] A. Sommerfeld, *Über verzweigte Potentiale im Raum*, Proc. London Math. Soc. (1) **28**, 395–429 (1897).

[5] Basset, Proc. London Math. Soc. (1) **16**, 286 (1885).

VIBRATIONAL PROBLEMS IN ELLIPTICAL COORDINATES*

BY

N. W. McLACHLAN

(*University of London*)

1. Introduction. We commence with the differential equation

$$\frac{\partial^2\zeta_0}{\partial x^2}+\frac{\partial^2\zeta_0}{\partial y^2}-K\frac{\partial^2\zeta_0}{\partial t^2}=0, \tag{1.1}$$

K being a constant. Writing $\zeta_0=e^{i\omega t}\zeta(x, y)$ in (1.1), it becomes

$$\frac{\partial^2\zeta}{\partial x^2}+\frac{\partial^2\zeta}{\partial y^2}+k_1^2\zeta=0, \tag{1.2}$$

with $k_1^2=\omega^2K$. This is the familiar two-dimensional wave equation for sinusoidal time-displacement, expressed in rectangular coordinates. The ordinary differential equations used herein, into which (1.2), expressed in elliptical coordinates, may be decomposed, are identical in form with those obtained by Mathieu[1] when he solved the problem of the elliptical membrane. Their preferred canonical forms are

$$\frac{d^2y}{dz^2}+(a-2q\cos 2z)y=0, \tag{1.3}$$

and

$$\frac{d^2y}{dz^2}-(a-2q\cosh 2z)y=0, \tag{1.4}$$

with $q>0$. In dealing with an elliptical plate, these equations are used with $q\gtrless 0$. (1.3) is known as the ordinary Mathieu equation, and (1.4) as the *modified* Mathieu equation. They are derivable from each other by the substitution $\pm iz$ for z. Recently lists of solutions of (1.3), (1.4), and the corresponding equations for $q<0$, have been published.[2,3,4] By aid of these we shall derive formal solutions pertaining to the vibrational modes of:

(a) A uniform, homogeneous, loss-free stretched membrane in the form of an elliptical ring;[5]

(b) Water in a lake of uniform depth whose plan view is an elliptical ring;

(c) A uniform, homogeneous, loss-free, elastic elliptical plate;

* Received Oct. 28, 1946.

[1] E. Mathieu, *Mémoire sur le mouvement vibratoire d'une membrane de forme elliptique*, Jour. de Math. Pures et Appliquées. **13,** 137 (1868).

[2] W. G. Bickley, and N. W. McLachlan, *Mathieu functions of integral order and their tabulation*, MTAC, **2,** 1, 1946.

[3] N. W. McLachlan, *Mathieu functions and their classification*, Jour. Math. Phys., **25,** 209, 1946; *Theory and application of Mathieu functions*, Oxford Press, New York, 1947.

[4] J. A. Stratton, P. M. Morse, L. J. Chu, and A. Hutner, *Elliptic cylinder and spheroidal wave functions*, Wiley, & Sons New York, 1942.

[5] An elliptical ring is that portion of the ellipse enclosed by the inner and outer confocal bounding ellipses.

(d) A uniform, homogeneous, loss-free, elastic elliptical ring plate.

In all cases, as the outer bounding ellipse tends to a circle, the formulae degenerate to those already known for a circular boundary.

2. Elliptical ring membrane. In this case $k_1^2=\omega^2\rho_1/\tau$, where $\rho_1=$mass of membrane per unit area, $\tau=$radial tension per unit arc length,[6] both of the latter being constant. To derive the appropriate formal solutions, we first transform (1.1) to elliptical coordinates,[7] where

$$x = h\cosh\xi\cos\eta, \qquad y = h\sinh\xi\sin\eta, \tag{2.1}$$

$2h$ being the interfocal distance. Then if ζ is the displacement of a point (ξ, η) on membrane, we get

$$\frac{\partial^2\zeta}{\partial\xi^2}+\frac{\partial^2\zeta}{\partial\eta^2}+\tfrac{1}{2}k^2(\cosh 2\xi-\cos 2\eta)\zeta = 0. \tag{2.2}$$

with $k^2=\frac{1}{4}k_1^2h^2=\omega^2h^2\rho_1/4\tau$.

3. Solutions of (2.2). The appropriate type take the form

$$\zeta(\xi, \eta) = \chi(\xi)\psi(\eta) \tag{3.1}$$

where ψ and χ are functions of η, ξ alone, respectively, which satisfy the ordinary equations

$$\frac{d^2\psi}{d\eta^2}+(a-2q\cos 2\eta)\psi = 0, \tag{3.2}$$

and

$$\frac{d^2\chi}{d\xi^2}-(a-2q\cosh 2\xi)\chi = 0. \tag{3.3}$$

a is the arbitrary separation constant, while $q=k^2$, $k>0$. These equations have the same forms as (1.3), (1.4), and we designate their first and second linearly independent solutions by

$$\psi_1(\eta, q), \qquad \psi_2(\eta, q); \tag{3.4}$$

$$\chi_1(\xi, q), \qquad \chi_2(\xi, q). \tag{3.5}$$

As shown in the papers mentioned in Footnotes 2, 3, 4 these solutions have a plurality of forms, so the next step is to select those appropriate to the problem of the ring membrane.

4. Physical conditions. Consider the maximum displacement of all points on a confocal ellipse on the membrane. If we start at $\eta=0$ and move round the ellipse counterclockwise, the maximum displacement varies continuously. If *may* be repeated between $\eta=\pi$ and $\eta=2\pi$, but it is *always* repeated at interval 2π. Thus ζ is single-valued and periodic in the coordinate η, the period being either π or 2π, ac-

[6] To obtain uniform tension, the membrane may be stretched with uniform radial tension over a circular frame, the elliptical rings then being clamped in position within the frame. The plane of the membrane, when at rest, is horizontal.

[7] A simple method of transformation is exemplified in Phil. Mag. **36**, 600, (1945).

cording to the mode of vibration. Consequently $\psi_1(\eta)$ and $\psi_2(\eta)$ must be functions such that either

$$\psi(\eta, q) = \psi(\eta + \pi, q), \tag{4.1}$$

or

$$\psi(\eta, q) = \psi(\eta + 2\pi, q). \tag{4.2}$$

According to the references in footnotes 2, 3, the only functions which satisfy these conditions are, respectively,

$$\psi_1(\eta, q) = ce_{2n}(\eta, q), \quad se_{2n+2}(\eta, q), \quad [a_{2n}, b_{2n+2}],^{8} \tag{4.3}$$

and

$$\psi_1(\eta, q) = ce_{2n+1}(\eta, q), \quad se_{2n+1}(\eta, q), \quad [a_{2n+1}, b_{2n+1}],^{8} \tag{4.4}$$

$n = 0, 1, 2, \cdots$ The functions in (4.3), (4.4) are first solutions of (3.2). Linearly independent second solutions are *non-periodic* and, therefore, inadmissible here. There is no physical reason for discriminating between independent solutions of (3.3) so we may include both. Then,[2,3]

$$\chi_1(\xi, q) = Ce_m(\xi, q), \quad Se_m(\xi, q), \quad [a_m, b_m], \tag{4.5}$$

and

$$\chi_2(\xi, q) = Fey_m(\xi, q), \quad Gey_m(\xi, q), \quad [a_m, b_m], \tag{4.6}$$

where $m = 2n$, $2n+1$, $2n+2$, as the case may be. The second alternative solutions $Fe_m(\xi, q)$, $Ge_m(\xi, q)$ may be used for $\chi_2(\xi, q)$, but for reasons stated in the references mentioned in Footnote 3, the solutions (4.6) are preferable.

5. The formal solution. This is to be constructed from (3.1), and (4.3)–(4.6). We arrange the $\chi\psi$ in groups, the functions in each group having the same characteristic number. Thus, introducing the real part of the time factor from Sec. 1 with an arbitrary phase angle, we have the component solutions of order m, corresponding to a_m and b_m, namely,

$$\zeta_m(\xi, \eta, t) = [C_m Ce_m(\xi, q) + F_m Fey_m(\xi, q)]ce_m(\eta, q) \cos(\omega_m t + \epsilon_m), \tag{5.1}$$

and

$$\bar{\zeta}_m(\xi, \eta, t) = [S_m Se_m(\xi, q) + G_m Gey_m(\xi, q)]se_m(\eta, q) \cos(\bar{\omega}_m t + \bar{\epsilon}_m). \tag{5.2}$$

In (5.1), (5.2), C_m, F_m, S_m, G_m, are arbitrary constants determinable from the displacement and velocity over the surface of the membranal ring at $t = 0$, while ω_m, $\bar{\omega}_m$ are the respective pulsatances of the mth free modes of vibration, and ϵ_m, $\bar{\epsilon}_m$ their relative phase angles. ω_m and $\bar{\omega}_m$ are different as will be evident from the next section. For *Ce*, *ce*, *Fey*, $m = 0, 1, 2, \cdots$, while for *Se*, *se*, *Gey*, $m = 1, 2, \cdots$. The complete solution, without the time factor, maybe written

$$\zeta(\xi, \eta) = \sum_{m=0}^{\infty} \zeta_m(\xi, \eta) + \sum_{m=1}^{\infty} \bar{\zeta}_m(\xi, \eta). \tag{5.3}$$

[8] These are the characteristic numbers corresponding to the Mathieu functions *ce*, *se* of integral order $2n$, $2n+1$, $2n+2$.

6. Vibrational modes of ring membrane. We use the boundary conditions $\xi=\xi_0$, $\zeta=0$ at the outer clamp, $\xi=\xi_1$, $\zeta=0$ at the inner clamp. For modes of order m corresponding to characteristic number a_m, by (5.1) at the outer clamp

$$[C_m Ce_m(\xi_0, q) + F_m Fey_m(\xi_0, q)]ce_m(\eta, q) = 0. \tag{6.1}$$

At the inner clamp

$$[C_m Ce_m(\xi_1, q) + F_m Fey_m(\xi_1, q)]ce_m(\eta, q) = 0. \tag{6.2}$$

These equations are independent of η at the clamps, so for C_m, F_m to be non-zero, we must have

$$Ce_m(\xi_0, q)Fey_m(\xi_1, q) - Ce_m(\xi_1, q)Fey_m(\xi_0, q) = 0, \qquad (m \geqq 0) \tag{6.3}$$

which is the pulsatance equation. Moreover, corresponding to each value of $q_{m,p}$, $p=1, 2, 3, \cdots$ for which (6.3) is satisfied, the elliptical ring has a vibrational mode of order m and rank p. Since $q_{m,p}=k^2_{m,p}=\omega^2_{m,p}h^2\rho_1/4\tau$, the pulsatances of the modes are given by

$$\omega^2_{m,p} = 4\tau q_{m,p}/\, h^2\rho_1, \tag{6.4}$$

$$= 4\tau q_{m,p}/a^2e^2\rho_1, \tag{6.5}$$

where $h=ae$, a being the semi-major axis, and e the eccentricity of the outer bounding ellipse. $q_{m,p}$ are the parametric zeros of (6.3), and determine the positions of the nodal ellipses, of which there are $(p-1)$ within the clamping rings. The nodal hyperbolae are determined for any $q=q_{m,p}$ by

$$ce_m(\eta, q) = 0, \qquad (m > 0). \tag{6.6}$$

For the set of modes corresponding to characteristic number b_m, the pulsatance equation defining the nodal ellipses is

$$Se_m(\xi_0, q)Gey_m(\xi_1, q) - Se_m(\xi_1, q)Gey_m(\xi_0, q) = 0, \qquad (m > 0), \tag{6.7}$$

where $q=\bar{q}_{m,p}$. The nodal hyperbolae, for any $q=\bar{q}_{m,p}$, are given by

$$se_m(\eta, q) = 0, \qquad (m > 0). \tag{6.8}$$

7. Elliptical ring lake. The formal analysis is identical with that for the ring membrane, but the condition at the inner and outer boundaries is zero velocity normal to them. Thus if ζ represents the tide height or vertical displacement of the water from its equilibrium level, we have[9]

$$\partial\zeta/\partial\xi = 0 \quad \text{at} \quad \xi = \xi_0, \quad \text{and also at} \quad \xi = \xi_1. \tag{7.1}$$

The conditional equations have the forms (6.1), (6.2), but the functions are replaced by their first derivatives with respect to ξ. Hence by (6.3) the pulsatance equation is

$$Ce_m(\xi_0, q)Fey_m'(\xi_1, q) - Ce_m'(\xi_1, q)Fey_m'(\xi_0, q) = 0, \qquad (m \geqq 0), \tag{7.2}$$

q having those values $q'_{m,p}$, $p=1, 2, \cdots$ for which (7.2) is satisfied. These determine the nodal ellipses, while the nodal hyperbolae for the same values of q are given by

[9] H. Jeffreys, *Free oscillations of water in an elliptical lake*, Proc. Lond. Math. Soc. **23**, 455 (1924).

(6.6). In the present problem $k^2=q=\omega^2h^2/gd$, g being the acceleration due to gravity, and d the uniform depth of the lake.

For the set of modes corresponding to characteristic number b_m, we have

$$Se_m'(\xi_0, q)Gey_m'(\xi_1, q) - Se_m'(\xi_1, q)Gey_m'(\xi_0, q) = 0, \qquad (m > 0), \tag{7.3}$$

and

$$se_m(\eta, q) = 0, \qquad (m > 0), \tag{7.4}$$

with $q=\bar{q}_{m,p}'$

8. Elliptical plate. Let ρ be the density, t the uniform thickness, $\sigma<1$ Poisson's ratio, E the modulus of elasticity, and $c^2=Et^2/12\rho(1-\sigma)$, whose dimensions are (velocity).[2] Then if ζ varies as $e^{i\omega t}$ in the various vibrational modes, $k_1^2=\omega^2/c^2$, and rotation effects are ignored, the differential equation of motion expressed in rectangular coordinates, may be shown to be

$$\frac{\partial^4\zeta}{\partial x^4}+\frac{\partial^4\zeta}{\partial y^4}+\frac{2\partial^4\zeta}{\partial x^2\partial y^2} - k_1^4\zeta = 0, \tag{8.1}$$

or

$$\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+k_1^2\right)\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}-k_1^2\right)\zeta = 0. \tag{8.2}$$

If a function ζ_1 satisfies

$$\frac{\partial^2\zeta_1}{\partial x^2}+\frac{\partial^2\zeta_1}{\partial y^2}+k_1^2\zeta_1 = 0, \tag{8.3}$$

and another function ζ_2 satisfies

$$\frac{\partial^2\zeta_2}{\partial x^2}+\frac{\partial^2\zeta_2}{\partial y^2}-k_1^2\zeta_2 = 0, \tag{8.4}$$

their sum, with the appropriate arbitrary constants, is a solution of (8.1), (8.2). To obtain this solution, (8.3), (8.4) are expressed in elliptical coordinates, thereby yielding two equations of the form (2.2), with $+k^2$ in one and $-k^2$ in the other. These may be split up into pairs of equations like (3.2), (3.3), one pair being associated with $+q$ and the other with $-q$. Then omitting the time factor, the solutions of order m corresponding to (8.3) are

$$\zeta_1^{(m)}(\xi, \eta) = [C_mCe_m(\xi, q) + F_mFey_m(\xi, q)]ce_m(\eta, q), \qquad (a = a_m), \tag{8.5}$$

$$\zeta_1^{(m)}(\xi, \eta) = [S_mSe_m(\xi, q) + G_mGey_m(\xi, q)]se_m(\eta, q), \qquad (a = b_m). \tag{8.6}$$

The solutions corresponding to (8.4) are

$$\zeta_2^{(m)}(\xi, \eta) = [\overline{C}_mCe_m(\xi, -q) + \overline{F}_mFek_m(\xi, -q)]ce_m(\eta, -q) \tag{8.7}$$

for $a=a_{2n}$, $m=2n$; $a=b_{2n+1}$, $m=2n+1$, and

$$\zeta_2^{(m)}(\xi, \eta) = [\overline{S}_mSe_m(\xi, -q) + \overline{G}_mGek_m(\xi, -q)]se_m(\eta, -q) \tag{8.8}$$

for $a=b_{2n}$, $m=2n$; $a=2n+1$, $m=2n+1$. In these cases $q<0$ in (2.2) and the functions $Fek_m(\xi, -q)$, $Gek_m(\xi, -q)$ have been used in preference to $Fey_m(\xi, -q)$, $Gey_m(\xi, -q)$, since the former are real, while the latter are complex if ξ is real.[3]

On the interfocal line of the ellipse $\xi=0$. In crossing this line orthogonally from $(0, \eta)$ to $(0, -\eta)$, we must have

(a) continuity of displacement, so $\zeta(0, \eta)=\zeta(0, -\eta)$, and

(b) continuity of gradient, so

$$\frac{\partial}{\partial\xi}[\zeta(\xi, \eta)]_{\xi\to 0} = -\frac{\partial}{\partial\xi}[\zeta(\xi, -\eta)]_{\xi\to 0}.$$

Remembering that $\zeta(\xi, \eta)=\chi(\xi)\psi(\eta)$, it may be demonstrated that with the product pairs Fey_mce_m, Gey_mse_m, Fek_mce_m, Gek_mse_m, conditions (a), (b) cannot *both* be satisfied. Hence in (8.5)–(8.8) we put $F_m=G_m=\overline{F}_m=\overline{G}_m=0$. Thus the complete formal solution is

$$\zeta(\xi, \eta) = \sum_{m=0}^{\infty} C_m Ce_m(\xi, q)ce_m(\eta, q) + \overline{C}_m Ce_m(\xi, -q)ce_m(\eta, -q) + \sum_{m=1}^{\infty} S_m Se_m(\xi, q)se_m(\eta, q) + \overline{S}_m Se_m(\xi, -q)se_m(\eta, -q). \tag{8.9}$$

9. Boundary conditions. When the periphery of the plate is clamped, $\zeta=\partial\zeta/\partial\xi=0$ at $\xi=\xi_0$. Thus for mode m corresponding to characteristic number a_{2n}, we have

$$C_{2n}Ce_{2n}(\xi_0, q)ce_{2n}(\eta, q) + \overline{C}_{2n}Ce_{2n}(\xi_0, -q)ce_{2n}(\eta, -q) = 0, \tag{9.1}$$

and

$$C_{2n}Ce'_{2n}(\xi_0, q)ce_{2n}(\eta, q) + \overline{C}_{2n}Ce'_{2n}(\xi_0, -q)ce_{2n}(\eta, -q) = 0. \tag{9.2}$$

For C_{2n}, $\overline{C}_{2n}$ to be non-zero, we must have

$$[Ce_{2n}(\xi_0, q)Ce'_{2n}(\xi_0, q) - Ce_{2n}(\xi_0, -q)Ce'_{2n}(\xi_0, q)]ce_{2n}(\eta, q)ce_{2n}(\eta, -q) = 0. \tag{9.3}$$

Hence the pulsatance equation is

$$Ce_{2n}(\xi_0, -q)Ce'_{2n}(\xi_0, q) - Ce'_{2n}(\xi_0, -q)Ce_{2n}(\xi_0, q) = 0, \qquad (n \geqq 0), \tag{9.4}$$

or

$$\frac{d}{d\xi}[Ce_{2n}(\xi, q)/Ce_{2n}(\xi, -q)]_{\xi=\xi_1} = 0. \tag{9.5}$$

This equation is satisfied when $q=q'_{2n,s}$ say, where $s=1, 2, 3, \cdots$, and these roots define a system of confocal nodal ellipses. For the same values of q, by (9.3) we have the equations whose roots define two systems of nodal hyperbolae, namely,

$$ce_{2n}(\eta, q) = 0, \quad \text{and} \quad ce_{2n}(\eta, -q) = 0 \qquad (n > 0). \tag{9.6}$$

For modes associated with the characteristic numbers a_{2n+1}, b_{2n+1}, we take the respective equations, with $m=(2n+1)$,

$$C_m Ce_m(\xi, q)ce_m(\eta, q) + \bar{S}_m Se_m(\xi, -q)se_m(\eta, -q) = 0, \tag{9.7}*$$

and

$$S_m Se_m(\xi, q)se_m(\eta, q) + \bar{C}_m Ce_m(\xi, -q)ce_m(\eta, -q) = 0. \tag{9.8}*$$

Using the boundary conditions in (9.7) yields for $a = a_{2n+1}$, the equations

$$\frac{d}{d\xi}\left[Ce_m(\xi, q)/Se_m(\xi, -q)\right]_{\xi=\xi_0} = 0, \tag{9.9}$$

and

$$ce_m(\eta, q) = 0, \qquad se_m(\eta, -q) = 0, \tag{9.10}$$

for $q = q'_{m,s}$, these being the roots of (9.9).

Similarly for (9.8) we derive for $a = b_{2n+1}$, the equations

$$\frac{d}{d\xi}\left[Se_m(\xi, q)/Ce_m(\xi, -q)\right]_{\xi=\xi_0} = 0, \tag{9.11}$$

and

$$se_m(\eta, q) = 0, \qquad ce_m(\eta, -q) = 0, \tag{9.12}$$

for $q = \bar{q}'_{m,s}$, these being the roots of (9.11).

For characteristic number $a = b_{2n}$, the equations may be obtained from (9.5), (9.6) by writing *Se* for *Ce*, *se* for *ce*, $\bar{q}'_{2,ns}$ for $q'_{2,ns}$, $n \geqq 1$.

10. Elliptical ring plate. When an elliptical plate is clamped at its periphery and also at an internal confocal ellipse, it becomes a ring. Defining the inner and outer boundaries by ξ_1 and ξ_0, the conditions to be satisfied are $\zeta = \partial\zeta/\partial\xi = 0$ at $\xi = \xi_1$ and ξ_0, where $0 \leqq \xi_1 < \xi_0$. If ξ_1 were zero, the inner clamp would be on the interfocal line of length $2h$. By Sec. 8, the formal solution for characteristic number a_{2n}, is

$$\begin{aligned}\zeta^{(2n)} = {} & \left[C_{2n}Ce_{2n}(\xi, q) + F_{2n}Fey_{2n}(\xi, q)\right]ce_{2n}(\eta, q) \\ & + \left[\bar{C}_{2n}Ce_{2n}(\xi, -q) + \bar{F}_{2n}Fek_{2n}(\xi, -q)\right]ce_{2n}(\eta, -q).\end{aligned} \tag{10.1}$$

Then at $\xi = \xi_0$ and at $\xi = \xi_1$, we must have

$$\begin{aligned}&\left[C_{2n}Ce_{2n}(\xi, q) + F_{2n}Fey_{2n}(\xi, q)\right]ce_{2n}(\eta, q) \\ &\qquad + \left[\bar{C}_{2n}Ce_{2n}(\xi, -q) + \bar{F}_{2n}Fek_{2n}(\xi, -q)\right]ce_{2n}(\eta, -q) = 0,\end{aligned} \tag{10.2}$$

and

$$\begin{aligned}&\left[C_{2n}Ce'_{2n}(\xi, q) + F_{2n}Fey'_{2n}(\xi, q)\right]ce_{2n}(\eta, q) \\ &\qquad + \left[\bar{C}_{2n}Ce'_{2n}(\xi, -q) + \bar{F}_{2n}Fek'_{2n}(\xi, -q)\right]ce_{2n}(\eta, -q) = 0.\end{aligned} \tag{10.3}$$

Thus there are four equations from which we can eliminate the four arbitrary constants, and so derive the pulsatance equation, and those for the nodal hyperbolae. Similar equations may be derived corresponding to a_{2n+1}, b_{2n+1}, b_{2n+2}.

11. Examples. 1°. *Elliptical ring membrane.* The first pulsatance equation is (6.3),

* It should be observed that $ce_{2n+1}(\eta, q)$ and $se_{2n+1}(\eta, -q)$ have $a = a_{2n+1}$, while for $se_{2n+1}(\eta, q)$ and $ce_{2n+1}(\eta, -q)$, $a = b_{2n+1}$. Similarly for *Ce* and *Se*.

and we have to find those values $q_{m,p}=\omega^2_{m,p}h^2\rho_1/4\tau$ which satisfy it. In the absence of tabular values of the two functions involved, calculation of the values q for the modes of lower order would be tedious. However, by aid of formulae asymptotic in $k=+q^{1/2}$, we can easily determine the approximate roots of (6.3). When $q=k^2$ is large enough[3]

$$\left.\begin{matrix} Ce \\ Fey \end{matrix}\right\}_m (\xi, q) \sim \frac{K_m}{(\cosh \xi)^{1/2}} \begin{matrix} \cos \\ \sin \end{matrix} (v_r - \theta_r), \tag{11.1}$$

where K_m is a constant dependent upon q, $v_r=2k \sinh \xi_r$, and $\theta_r=(2m+1) \tan^{-1} (\tanh \frac{1}{2}\xi_r)$. Using (11.1) in (6.3) leads to

$$\sin [(v_0 - v_1) - (\theta_0 - \theta_1)] = 0, \tag{11.2}$$

so

$$(v_0 - v_1) = p\pi + (\theta_0 - \theta_1), \tag{11.3}$$

p integral $\geqq 1$. Thus (11.3) gives

$$k_{m,p} \simeq (p\pi + \theta_0 - \theta_1)/2(\sinh \xi_0 - \sinh \xi_1). \tag{11.4}$$

If e is the eccentricity of a confocal ellipse ξ, $e^{-1}=\cosh \xi$ and, therefore, $\sinh \xi=e^{-1}(1-e^2)^{1/2}$. Substituting this into (11.4) yields

$$\begin{aligned} q_{m,p} &= k^2_{m,p} \\ &\simeq (p\pi + \theta_0 - \theta_1)^2 e_0^2 e_1^2/4[e_0^2 + e_1^2 - 2e_0^2 e_1^2 - 2e_1 e_0\{(1 - e_1^2)(1 - e_0^2)\}^{1/2}], \end{aligned} \tag{11.5}$$

e_0, e_1 being the eccentricities of the outer and inner bounding ellipses, respectively. The accuracy of (11.5) improves with increase in e and p. The second pulsatance equation (6.7) may be treated in a similar manner.

2°. *Elliptical plate.* Referring to (9.5), we shall consider the modes of order $2n$ corresponding to the characteristic number a_{2n}. As before, owing to absence of tabular values, we shall assume that k is large enough for (11.1) to be used. This formula applies when $q>0$, but we also require one for $q<0$. To derive this we write $(\xi+\frac{1}{2}\pi i)$ for ξ in (11.1), since this substitution changes the sign of q in (3.3) of which (11.1) is an approximate solution when q is large and positive. As explained in the paper quoted in Footnote 3, it is also necessary to multiply by $(-1)^n$ and select the real part of the formula, since ξ is real. Then if $|\xi| \gg \frac{1}{2}\pi$, we obtain

$$Ce_{2n}(\xi, -q) \sim \overline{K}_{2n}\chi_1/(\sinh \xi)^{1/2}, \tag{11.6}$$

where $\chi_1=\cosh u \cos \phi_{2n}+\sinh u \sin \phi_{2n}$, $u=2k \cosh \xi$, $\phi_{2n}\simeq(4n+1) \tan^{-1} (\tanh \frac{1}{2}\xi)$, and $\overline{K}_{2n}$ is a constant dependent upon q. Thus from (11.6) and (11.1), we get

$$\left(\frac{\overline{K}_{2n}}{K_{2n}}\right) \frac{Ce_{2n}(\xi, q)}{Ce_{2n}(\xi, -q)} \sim \frac{(\tanh \xi)^{1/2} \cos \chi}{\chi_1}, \tag{11.7}$$

with $\chi=2k \sinh \xi-(4n+1) \tan^{-1} (\tanh \frac{1}{2}\xi)$. Performing the differentiation indicated in (9.5), and equating to zero, leads after a little reduction to the pulsatance equation

$$\tan \chi = (1/\chi' \sinh 2\xi) - (\chi_1'/\chi_1\chi'), \qquad \xi = \xi_0. \tag{11.8}$$

If ξ_0 is large enough, $\chi_1'/\chi_1\chi' \simeq -1$, while χ' sinh $2\xi_0 \gg 1$. Then (11.8) reduces to

$$\tan \chi \simeq -1, \quad \text{or} \quad \chi \simeq (s - \tfrac{1}{4})\pi, \tag{11.9}$$

s being a positive integer such that χ is large enough for the above approximate asymptotic formulae to be valid. Thus (11.9) entails

$$2k \sinh \xi_0 \simeq (s - \tfrac{1}{4})\pi + (4n + 1) \tan^{-1} (\tanh \tfrac{1}{2}\xi_0). \tag{11.10}$$

If ξ_0 is such that $2k \sinh \xi_0 \simeq ke^{\xi_0}$, and $\tanh \frac{1}{2}\xi \simeq 1$, we obtain

$$q_{2n,s} = k^2_{2n,s} \simeq (s + n)^2\pi^2 e^{-2\xi_0}. \tag{11.11}$$

A GENERALIZATION OF THE FINITE FOURIER TRANSFORMATION AND APPLICATIONS*

BY
IDA ROETTINGER
Ann Arbor, Michigan

Introduction. The purpose of this paper is to generalize and unify the methods used by Doetsch [8]†, Kniess [12], Koschmieder [13] and others to solve certain boundary value problems by finite Fourier integral transformations. To give an idea of the general method to be developed a formal solution of the following boundary value problem is here given by means of a particular transformation.

$$Y_{tt}(x, t) = Y_{xx}(x, t) + x, \qquad 0 < x < \pi, \qquad 0 < t,$$
$$Y(x, 0+) = 0, \qquad Y_t(x, 0+) = 0,$$
$$Y(0+, t) = 0, \qquad Y_x(\pi -, t) + hY(\pi -, t) = 0, \qquad h \neq 0.$$

Let $S\{F(x)\} = \int_0^\pi F(x) \sin k_n x dx = f_s(k_n)$, (compare section 1 below), where $\sin k_n x$ are the characteristic functions of $y''(x) + k^2 y(x) = 0$, $y(0) = 0$, $y'(\pi) + hy(\pi) = 0$, that is k_n, $n = 1, 2, \cdots$, are the roots of $\tan k\pi = -(k/h)$, $k > 0$, then, if $F(x)$, $F'(x)$ are continuous and $F''(x)$ is sectionally continuous in $(0, \pi)$

$$S\{F''(x)\} = -k_n^2 f_s(k_n) + k_n F(0) + \sin k_n \pi [F'(\pi) + hF(\pi)],$$

(see theorem 1 below).

This last formula applied to the above problem with respect to x yields the following transformed problem:

$$\frac{d^2 y_s(k_n, t)}{dt^2} = -k_n^2 y_s(k_n, t) + S\{x\},$$
$$y_s(k_n, 0+) = 0, \qquad dy_s(k_n, 0+)/dt = 0, \quad \text{where}$$
$$S\{x\} = (\sin k_n \pi) k_n^{-2} - (\pi \cos k_n \pi) k_n^{-1}.$$

The solution of the transformed problem is

$$y_s(k_n, t) = [(1 - \cos k_n t)/k_n^2] S\{x\}.$$

The inverse transformation $S^{-1}\{f_s(k_n)\}$ is given in terms of a Sturm-Liouville series (see section 2 below) as follows

$$S^{-1}\{f_s(k_n)\} = \sum_1^\infty N(k_n) f_s(k_n) \sin k_n x \quad \text{in} \quad (0, \pi),$$

where $N(k_n)$ denotes the normalization factor of the characteristic functions $\sin k_n x$, $n = 1, 2, \cdots$, .

* Received Oct. 18, 1946. Presented to the American Mathematical Society April 26, 1946. See also abstracts in Bull. Am. Math. Soc. vols. 50 and 51. The author expresses her gratitude to Professor Ruel V. Churchill for his advice and encouragement in the preparation of this paper. The content of this paper is part of a dissertation submitted in partial fulfillment of the requirements for the degree of doctor of philosophy in the University of Michigan.

† The numbers in brackets refer to the bibliography.

Thus the formal solution of the above boundary value problem is given by

$$Y(x, t) = \sum_{1}^{\infty} N(k_n)\left[(\sin k_n\pi)k_n^{-4} - (\pi \cos k_n\pi)k_n^{-3}\right](1 - \cos k_n t) \sin k_n x,$$

The type of problems to which the method is applicable is described in the summary (see Sec. 11 below).

Since there are at present no readily applicable general existence theorems for solutions of the class of boundary value problems considered here, the existence and uniqueness of the solution should be established in each particular case. However, this procedure is not carried through in the problems solved in this paper, since the main purpose here is to set up a method which leads quickly to a formal solution. The problems solved are chosen merely to give an illustration of the method.

The method can be compared with that of the Laplace transformation (see remark in Sec. 3 below). As in the case of the Laplace transformation the present operational method does not claim to solve problems which cannot be solved by any other method. Its advantage lies, just as in the case of the Laplace transformation, in its direct, short and systematic approach. Problems in partial differential equations which by a suitable change of variables are brought into a form in which variables can be separated, can be solved directly. Tables of transforms, which are, except for normalization factors, tables of generalized Fourier coefficients, save time in the computation of solutions of practical problems. Furthermore this operational method has, due to certain theorems the advantage of systematically finding *closed form solutions* (see Secs. 6, 9, 10, 11 below) and thus exhibiting qualitative properties of a solution which may not readily be found by the usual methods.

1. **Definitions of the transformations** S, C **and** T. *Definition* 1. Let $K = \{k\}$ be a set of real numbers and let $F(x)$ be a sectionally continuous function in $(0, \pi)$. The transformations $S\{F(x)\}$ and $C\{F(x)\}$ are defined by the equations

$$S\{F(x)\} = \int_0^{\pi} F(x) \sin kx dx = f_s(k) \tag{1}$$

$$C\{F(x)\} = \int_0^{\pi} F(x) \cos kx dx = f_c(k) \tag{2}$$

respectively.

The transformations (1) and (2) are called the general finite Fourier sine and cosine transformation respectively, relative to the interval $(0, \pi)$ and the set K. The transformations (1) and (2) map the class of functions $F(x)$ onto a certain class of functions $f(k)$.

A Sturm-Liouville problem. The characteristic functions of the following Sturm-Liouville problem are used in the definition of the transformation T below.

$$\left.\begin{aligned} &y''(x) + k^2 y(x) = 0, \quad \text{in} \quad (0, \pi), \\ &L_1(y) = a_1 y(0) + a_2 y'(0) + a_3 y(\pi) + a_4 y'(\pi) = 0, \\ &L_2(y) = b_1 y(0) + b_2 y'(0) + b_3 y(\pi) + b_4 y'(\pi) = 0, \end{aligned}\right\} \tag{3}$$

also with specializations on part of the constant coefficients a_i, b_i, $i = 1, 2, 3, 4$. It is assumed that the L_j, $j = 1, 2$, are linearly independent and that

$$a_1b_2 - a_2b_1 = a_3b_4 - a_4b_3. \tag{4}$$

This last condition is to guarantee real characteristic values (see for instance [10] vol. 1 p. 352).

The characteristic functions $\varphi_{k_n}(x)$, $n=1, 2, \cdots$, of (3) are given by

$$\varphi_{k_n}(x) = A \sin k_n x + B \cos k_n x, \tag{5}$$

where

$$\begin{aligned}(A/B) &= -(a_1 + a_3 \cos k_n\pi - a_4k_n \sin k_n\pi)/(k_na_2 + a_3 \sin k_n\pi + a_4 k_n \cos k_n\pi)\\ &= -(b_1 + b_3 \cos k_n\pi - b_4k_n \sin k_n\pi)/(k_nb_2 + b_3 \sin k_n\pi + b_4k_n \cos k_n\pi)\end{aligned}$$

and k_n are the roots of

$$D(k) \equiv \begin{vmatrix} (ka_2 + a_3 \sin k\pi + a_4k \cos k\pi) & (a_1 + a_3 \cos k\pi - a_4k \sin k\pi) \\ (kb_2 + b_3 \sin k\pi + b_4k \cos k\pi) & (b_1 + b_3 \cos k\pi - b_4k \sin k\pi) \end{vmatrix} = 0. \tag{6}$$

By use of (4), Eq. (6) reduces to

$$\begin{aligned}2k(a_1b_2 - a_2b_1) = &[(a_3b_1 - a_1b_3) + k^2(a_4b_2 - a_2b_4)] \sin k\pi \\ &+ [(a_4b_1 - a_1b_4) + (a_2b_3 - a_3b_2)]k \cos k\pi.\end{aligned} \tag{7}$$

Remarks concerning $D(k)$. (Compare [10] vol. 1, p. 362.) If $D(k)=0$ has no roots, then (3) has no solution.

If and only if $D(k)$ is of rank 0, that is, if every term in $D(k)$ is 0, then the characteristic values are called double, since in this case (3) has two linearly independent characteristic functions with the same characteristic values, e.g. $a_1=b_2=-a_3-b_4=1$ all other a_i, b_i zero. In this case the characteristic values are $2n$, $n=1, 2, \cdots$, and $\sin 2nx$ as well as $\cos 2nx$ are characteristic functions of (3).

If $D(k)$ is of rank greater than zero the characteristic values are single. This is the case in particular in the first, second and third boundary value problem.

Definition 2. Let $K=\{k_n\}$, $n=1, 2, \cdots$, be the sequence of characteristic values of (3), that is, the roots of (7). Let

$$\varphi_{k_n}(x) = A_{k_n} \sin k_n x + B_{k_n} \cos k_n x \tag{8}$$

be given by (5), where

$$\begin{aligned}A_{k_n} &= -a_1 + a_3 \cos k_n\pi - a_4k_n \sin k_n\pi,\\ B_{k_n} &= -(k_na_2 + a_3 \sin k_n\pi + a_4k_n \cos k_n\pi).\end{aligned}$$

And let $F(x)$ be a sectionally continuous function in $(0, \pi)$. The transformation $T\{F(x)\}$ is defined by the equation

$$T\{F(x)\} = \int_0^\pi F(x)\varphi_{k_n}(x)dx = f(k_n). \tag{9}$$

The transformations S, C and T are linear. The interval $(0, \pi)$ is chosen for convenience and without loss of generality. The transformation (9) maps the class of functions $F(x)$ onto a class of sequences of numbers; except for a normalization factor, each sequence is the set of Sturm-Liouville coefficients of the corresponding $F(x)$ in terms of the characteristic functions $\varphi_{k_n}(x)$. The restriction to sectionally continuous

functions was made in view of applications. In general the functions $F(x)$ need only belong to $L^2(0, \pi)$: i.e. need only be Lebesgue square integrable.

2. Inverse transformations. Since the function $\varphi_{k_n}(x)$ form a complete set of orthogonal functions on the interval $(0, \pi)$, the transformation (9) has an inverse transformation in the form of a Sturm-Liouville expansion. For, let $F(x)$ be sectionally continuous and $[F'(x)]^2$ integrable in $(0, \pi)$ and at a point of discontinuity x_0, let $F(x)$ be defined as $F(x_0)=\frac{1}{2}[F(x_0+0)+F(x_0-0)]$, $0<x_0<\pi$, and let $N(k_n)$ denote the normalization factor of the functions (8), then the expansion in the characteristic functions (8) of the function $F(x)$ converges to the function $F(x)$ in $(0, \pi)$; i.e.,

$$F(x) = \sum_1^\infty N(k_n)f(k_n)\varphi_{k_n}(x), \quad \text{in} \quad (0, \pi), \tag{10}$$

and the convergence is uniform and absolute in every closed subinterval of $(0, \pi)$ which does not contain a discontinuity (see [6] vol. 1, p. 371, compare also [5] pp. 268, 272). Equation (10) gives a formula for the inverse $T^{-1}\{f(k_n)\}$, the function whose T-transform is $f(k_n)$. The inverse is unique. Thus

$$T^{-1}\{f(k_n)\} = \sum_1^\infty N(k_n)f(k_n)\varphi_{k_n}(x), \quad \text{in} \quad (0, \pi), \tag{11}$$

where $N(k_n)$ is given by

$$\begin{aligned} 1/N(k_n) = {} & \tfrac{1}{2}[\pi(a_1^2 + a_3^2) + k_n^2\pi(a_2^2 + a_4^2) - a_1a_2 + a_3a_4] \\ & + \sin k_n\pi[k_n\pi(a_2a_3 - a_1a_4) + (1/k_n)(k_n^2a_2a_4 - a_1a_3)] \\ & + (\sin 2k_n\pi/4k_n)[k_n^2(a_4^2 + a_2^2) - (a_1^2 + a_3^2)] \\ & + \pi(k_n^2a_2a_4 + a_1a_3)\cos k_n\pi + (\cos 2k_n\pi/2)(a_1a_2 - a_3a_4). \end{aligned}$$

Similarly the inverse transformations $S^{-1}\{f_s(k)\}$ and $C^{-1}\{f_c(k)\}$ of the transformations $S\{F(x)\}$ and $C\{F(x)\}$ respectively are given in terms of a sine and cosine series respectively if K is the sequence of characteristic values of certain special cases of (3). e.g. $a_1\neq 0$, $b_3\neq 0$ all other a_i, b_i, zero, then $K=\{n\}$, $n=1, 2, \cdots$, and $S\{F(x)\}$ is the finite Fourier sine transformation as defined by Kniess (see [12]) and $S^{-1}\{f_s(n)\}$ is given by a Fourier sine series.

3. Transformation of derivatives. The following two lemmas can be proved by integration by parts.

Lemma 1. If $F(x)$, $F'(x)$, $\cdots$ $F^{(2r-1)}(x)$, are continuous and $F^{(2r)}(x)$ is sectionally continuous in $(0, \pi)$, $r=1, 2, \cdots$, then

$$\begin{aligned} S\{F^{(2r)}(x)\} = {} & (-1)^r k^{2r} S\{F(x)\} + \sum_{s=0}^{s=r-1} (-1)^s k^{2s}[k\{F^{(2r-2s-2)}(0) \\ & - F^{(2r-2s-2)}(\pi)\cos k\pi\} + F^{(2r-2s-1)}(\pi)\sin k\pi] \end{aligned} \tag{12}$$

and

$$\begin{aligned} C\{F^{(2r)}(x)\} = {} & (-1)^r k^{2r} C\{F(x)\} + \sum_{s=0}^{s=r-1} (-1)^s k^{2s}[kF^{(2r-2s-2)}(\pi)\sin k\pi \\ & + F^{(2r-2s-1)}(\pi)\cos k\pi - F^{(2r-2s-1)}(0)]. \end{aligned} \tag{13}$$

Lemma 2. If $F(x)$, $F'(x)$, $\cdots$ $F^{(2r-2)}(x)$, are continuous and $F^{(2r-1)}(x)$ is sectionally continuous in $(0, \pi)$, $r=1, 2, \cdots$, then

$$S\{F^{(2r-1)}(x)\} = (-1)^r k^{2r-1} C\{F(x)\} + \sin k\pi \sum_{s=0}^{s=r-1} (-1)^s k^{2s} F^{(2r-2s-2)}(\pi) + \sum_{s=1}^{s=r-1} (-1)^s k^{2s-1} [F^{(2r-2s-1)}(\pi) \cos k\pi - F^{(2r-2s-1)}(0)] \tag{14}$$

and

$$C\{F^{(2r-1)}(x)\} = (-1)^{r+1} k^{2r-1} S\{F(x)\} + \sin k\pi \sum_{s=1}^{s=r-1} (-1)^{s+1} k^{2s-1} F^{(2r-2s-1)}(\pi) + \sum_{s=0}^{s=r-1} (-1)^s k^{2s} [F^{(2r-2s-2)}(\pi) \cos k\pi - F^{(2r-2s-2)}(0)]. \tag{15}$$

Theorem 1. If $F(x)$, $F'(x)$, $\cdots$ $F^{(2r-1)}(x)$, are continuous and $F^{(2r)}(x)$ is sectionally continuous in $(0, \pi)$, $r=1, 2, \cdots$, and if $u_1=\varphi_{k_n}(0)$, $u_2=\varphi'_{k_n}(0)$, $u_3=\varphi_{k_n}(\pi)$, $u_4=\varphi'_{k_n}(\pi)$, ($\varphi_{k_n}(x)$ as given by (8)), then there exist numbers λ and μ such that

$$T\{F^{(2r)}(x)\} = (-1)^r k_n^{2r} T\{F(x)\} + \sum_{s=0}^{s=r-1} (-1)^s k_n^{2s} [\lambda L_1(F^{(2r-2s-2)}) + \mu L_2(F^{(2r-2s-2)})], \tag{16}$$

where

$$\lambda = (u_2 b_2 + u_1 b_1)/(a_1 b_2 - a_2 b_1) = k_n + (a_1 b_2 - a_2 b_1)^{-1} [(a_3 b_2 - a_4 b_1) k_n \cos k_n\pi - (a_3 b_1 + k_n^2 a_4 b_2) \sin k_n\pi]$$

and

$$\mu = (u_3 a_3 + u_4 a_4)/(a_1 b_2 - a_2 b_1) = (a_1 b_2 - a_2 b_1)^{-1} [(a_1 a_4 - a_2 a_3) k_n \cos k_n\pi + (a_1 a_3 + k_n^2 a_2 a_4) \sin k_n\pi].$$

Remark. Property (16) exhibits the usefulness of the transformation T in obtaining quickly a formal solution of the following type of boundary value problem:

$$\sum_{r=0}^{m} A_{2r} \frac{d^{2r} F(x)}{dx^{2r}} = G(x), \quad \text{in } (0, \pi), \quad \text{where}$$

$G(x)$, the constants A_{2r} and the quantities $L_j(F^{(2r-2s-2)})$, $j=1, 2$, $s=1, 2, \cdots r-1$, are assigned. The constants a_i, b_i, $i=1, 2, 3, 4$, are allowed to be different for each s. The transformation is also useful in boundary value problems in partial differential equations, where one or more variables behave like x in the above problem. Property (16) is analogous to the property of the Laplace transformation which expresses the transform of the nth derivative of a function in terms of the transform of the function and the value of the 0th, $\cdots$ $(n-1)$th derivative of the function at 0 (see for instance [5] p. 8).

Proof of theorem 1. By combining (12) and (13) $T\{F^{(2r)}(x)\}$ can be expressed as follows:

$$T\{F^{(2r)}(x)\} = (-1)^r k_n^{2r} T\{F(x)\} + \sum_{s=0}^{s=r-1} (-1)^s k_n^{2s} [u_2 F^{(2r-2s-2)}(0) - u_1 F^{(2r-2s-1)}(0) - u_4 F^{(2r-2s-2)}(\pi) + u_3 F^{(2r-2s-1)}(\pi)].$$

If

$$F^{(2r-2s-2)}(0) = w_1, \qquad F^{(2r-2s-1)}(0) = w_2,$$
$$F^{(2r-2s-2)}(\pi) = w_3, \qquad F^{(2r-2s-1)}(\pi) = w_4,$$

it remains to be shown that

$$u_2 w_1 - u_1 w_2 + u_3 w_4 - u_4 w_3 = \lambda \sum_1^4 a_i w_i + \mu \sum_1^4 b_i w_i,$$

which means it has to be shown that the following four equations are consistent:

$$u_2 = \lambda a_1 + \mu b_1$$
$$-u_1 = \lambda a_2 + \mu b_2$$
$$-u_4 = \lambda a_3 + \mu b_3$$
$$u_3 = \lambda a_4 + \mu b_4.$$

It can be left to the reader to obtain for instance λ from the first two equations and μ from the last two equations and to show that with these values of λ and μ the above equations are consistent.

4. Illustration of the operational method. Particular cases of transformations. As an illustration of the use of the operator T a formal solution of the following boundary value problem is to be established.

$$\Psi_{xx} + \Psi_{yy} = Q(x), \qquad 0 < x < \pi, \qquad 0 < y,$$

$\Psi(x, y)$ is to remain bounded as y approaches infinity and $\Psi(x, 0+) = 0$, furthermore, $a\Psi(0+, y) + b\Psi_x(0+, y) = E$, where E is a constant, and $c\Psi(\pi-, y) + d\Psi_x(\pi-, y) = 0$, where a, b, c, d are given non-zero constants such that $ad - bc \neq 0$.

This problem calls for the operator (9), where $\varphi_{k_n}(x)$ are the characteristic functions of $y''(x) + k^2 y(x) = 0$, $ay(0) + by'(0) = 0$, $cy(\pi) + dy'(\pi) = 0$. From (7) follows that the characteristic values k_n are the roots of $\tan k\pi = k(bc - ad)/(ac + k^2 bd)$, $k > 0$. From (9) it follows that

$$T\{F(x)\} = \int_0^\pi F(x)(a \sin k_n x - b k_n \cos k_n x) dx = f(k_n),$$

$n = 1, 2, \cdots$, and (11) yields

$$T^{-1}\{f(k_n)\} = \sum_1^\infty N(k_n) f(k_n)(a \sin k_n x - b k_n \cos k_n x), \quad \text{in} \quad (0, \pi), \tag{17}$$

where $N(k_n)$ is given by

$$1/N(k_n) = \tfrac{1}{2}(\pi a^2 + \pi k_n^2 b^2 - ab) + [(k_n^2 b^2 - a^2)/4k_n] \sin 2k_n\pi + (ab/2) \cos 2k_n\pi.$$

Equation (16) with $r=1$ yields

$$T\{F''(x)\} = -k_n^2 f(k_n) + k_n[aF(0) + bF'(0)] - [k_n^2 b/(c \sin k_n\pi + k_n d \cos k_n\pi)][cF(\pi) + dF'(\pi)]. \tag{18}$$

Formula (18) applied to the above problem with respect to x leads to the following transformed problem:

$$\psi_{yy}(k_n, y) - k_n^2\psi(k_n, y) + k_nE - q(k_n) = 0,$$

$\psi(k_n, y)$ remains bounded as $y\to\infty$ and $\psi(k_n, 0+)=0$, where ψ and q stand for the T-transforms of Ψ and Q respectively.

The solution of the transformed problem, that is the coefficient in the Sturm-Liouville expansion of the solution is

$$\psi(k_n, y) = [1 - \exp(-k_n y)][k_nE - q(k_n)]/k_n^2$$

and according to (17) the formal solution of the given problem is

$$\Psi(x, y) = \sum_1^\infty N(k_n)\{[1 - \exp(-k_n y)][k_nE - q(k_n)]/k_n^2\}(a \sin k_n x - bk_n \cos k_n x).$$

In the following sections a special case of the transformations S and C will be needed (for additional special cases see section 12). The kernels $\sin kx$ and $\cos kx$ respectively of the transformations are the characteristic functions of $y''(x)+k^2y(x)=0$ $y(0)=0$, $y'(\pi)=0$; and $y'(0)=0$, $y(\pi)=0$ respectively.
In these two cases $K=\{n-\frac{1}{2}\}$, $n=1, 2, \cdots$. The transformations are

$$S\{F(x)\} = \int_0^\pi F(x) \sin(n-\tfrac{1}{2})x\,dx = f_s(n-\tfrac{1}{2}), \qquad n = 1, 2, \cdots \tag{19}$$

$$C\{F(x)\} = \int_0^\pi F(x) \cos(n-\tfrac{1}{2})x\,dx = f_c(n-\tfrac{1}{2}), \qquad n = 1, 2, \cdots. \tag{20}$$

The inverse transformations are given by

$$S^{-1}\{f_s(n-\tfrac{1}{2})\} = \frac{2}{\pi}\sum_1^\infty f_s(n-\tfrac{1}{2}) \sin(n-\tfrac{1}{2})x \quad \text{in} \quad (0, \pi) \tag{21}$$

$$C^{-1}\{f_c(n-\tfrac{1}{2})\} = \frac{2}{\pi}\sum_1^\infty f_c(n-\tfrac{1}{2}) \cos(n-\tfrac{1}{2})x \quad \text{in} \quad (0, \pi). \tag{22}$$

And for the transform of an even derivative (12) and (13) yield

$$S\{F^{(2r)}(x)\} = -(n-\tfrac{1}{2})^{2r}f_s(n-\tfrac{1}{2}) + \sum_{s=0}^{s=r-1}(-1)^s(n-\tfrac{1}{2})^{2s}[(-1)^{n+1}F^{(2r-2s-1)}(\pi) + (n-\tfrac{1}{2})F^{(2r-2s-1)}(0)] \tag{23}$$

$$C\{F^{(2r)}(x)\} = -(n-\tfrac{1}{2})^{2r}f_c(n-\tfrac{1}{2}) + \sum_{s=0}^{s=r-1}(-1)^{s+1}(n-\tfrac{1}{2})^{2s}[F^{(2r-2s-1)}(0) + (n-\tfrac{1}{2})(-1)^nF^{(2r-2s-2)}(\pi)]. \tag{24}$$

Remark. In the above listed cases $D(k)$ is of rank greater than zero. If the rank of $D(k)$ is zero it merely has to be kept in mind that the sine as well as the cosine coefficients in the Sturm-Liouville expansion of the solution have to be found. An illustration of the operational method in this case is given in section 10.

5. Some properties of the transformations S and C when $K=\{n-\frac{1}{2}\}$, $n=1, 2, \cdots$. Theorems 2 to 5 below are analogous to theorems which holds for the finite Fourier sine and cosine transformations (compare [5] sections 95 and 96 also [19] chap. II sect. 2.1). and can be proved in a similar way.

Definition 3. Let the function $F(x)$ be defined in $(0, p)$. By the odd antiperiodic extension $F_1(x)$ of $F(x)$ with antiperiod $2p$ is meant $F_1(x)=F(x)$, in $(0, p)$; $F_1(-x) = -F_1(x)$, $F_1(x+2p)=-F_1(x)$ in $(-\infty, \infty)$.

Definition 4. Let the function $F(x)$ be defined in $(0, p)$. By the even antiperiodic extension $F_2(x)$ of $F(x)$ with antiperiod $2p$ is meant $F_2(x)=F(x)$, in $(0, p)$; $F_2(-x) = F_2(x)$, $F_2(x+2p)=-F_2(x)$, in $(-\infty, \infty)$.

Theorem 2. If $F(x)$ is sectionally continuous in $(0, \pi)$ and if $F_1(x)$ is the odd antiperiodic extension of $F(x)$ with antiperiod 2π and if a is any constant, then

$$\sin (n-\tfrac{1}{2})aS\{F(x)\} = \tfrac{1}{2}C\{F_1(x+a) - F_1(x-a)\}.$$

Theorem 3. Under the same assumption as in theorem 2

$$\cos (n-\tfrac{1}{2})aS\{F(x)\} = \tfrac{1}{2}S\{F_1(x+a) + F_1(x-a)\}.$$

Theorem 4. If $F(x)$ is sectionally continuous in $(0, \pi)$ and if $F_2(x)$ is the even antiperiodic extension of $F(x)$ with antiperiod 2π and if a is any constant, then

$$\sin (n-\tfrac{1}{2})aC\{F(x)\} = \tfrac{1}{2}S\{F_2(x-a) - F_2(x+a)\}.$$

Theorem 5. Under the same assumption as in theorem 4

$$\cos (n-\tfrac{1}{2})aC\{F(x)\} = \tfrac{1}{2}C\{F_2(x-a) + F_2(x+a)\}.$$

Similar theorems in the case when K is the sequence of characteristic values of (3) and for more general K can be proved by the use of almost periodic functions (see [15]).

6. Vibration of a horizontal string with one end fixed and one end sliding. Compare [5] sections 98, 99.) Let the end $x=0$ of a string be fixed and the end $x=\pi$ be looped about a vertical support along the line $x=\pi$. If a constant upward force acts on the loop, the displacements $Y(x, t)$ as the string is released from the position $Y=0$ satisfy the conditions:

$$\begin{aligned} Y_{tt}(x, t) &= a^2 Y_{xx}(x, t) + g, \qquad 0 < x < \pi, \qquad 0 < t, \\ Y(x, 0+) &= 0 \qquad Y_t(x, 0+) = 0, \\ Y(0+, t) &= 0 \qquad Y_x(\pi-, t) = -b, \end{aligned}$$

where g is the acceleration of gravity and b is the magnitude of the vertical force divided by the tension.

The S-transformation (19) applied with respect to x gives, according to (23), the following transformed problem:

$$\frac{d^2y(n-\frac{1}{2},t)}{dt^2}+(n-\tfrac{1}{2})^2a^2y(n-\tfrac{1}{2},t)+(-1)^{n+1}a^2b-gS\{1\}=0$$

$$y(n-\tfrac{1}{2},0+)=0, \qquad y_t(n-\tfrac{1}{2},0+)=0.$$

The solution of this transformed problem is

$$y(n-\tfrac{1}{2},t)=[1-\cos(n-\tfrac{1}{2})at]\{[g/a^2(n-\tfrac{1}{2})^3]+[b(-1)^n/(n-\tfrac{1}{2})^2]\}.$$

From the table in section 12 below it follows that

$$S^{-1}\{[b(-1)^n/(n-\tfrac{1}{2})^2]+[g/a^2(n-\tfrac{1}{2})^3]\}=-bx+[gx(2\pi-x)/2a^2]=F(x).$$

Using theorem 3 the solution of the given problem can be written in the form, $Y(x,t)=F(x)-\frac{1}{2}[F_1(x+at)+F_1(x-at)]$, where F_1 denotes the odd antiperiodic extension with antiperiod 2π of $F(x)$.

Similarly the problem with the end $x=0$ sliding and the end $x=\pi$ fixed can be solved by use of the C-transformation (20).

7. The convolution in the case $K=\{n-\frac{1}{2}\}$. The purpose of this section is to give formulas for the product of two transforms in terms of one transform. Kniess (see [12], compare also [19] 2.1) and Doetsch (see [8]) give such results for the finite Fourier sine and cosine transformations, i.e. when $K=\{n\}$, $n=1, 2, \cdots$. Here the analogous results are given for the S- and C-transformations when $K=\{n-\frac{1}{2}\}$ (compare also Koschmieder [13]). Analogous results when K is the set of characteristic values of certain Sturm-Liouville problems can be obtained by the use of almost period functions (see [15]). The proofs of theorems 6–9 are analogous to those of Kniess.

Definition 5. If $F(x)$ in $(-2\pi, 2\pi)$ and $G(x)$ in $(-\pi, \pi)$, are sectionally continuous, then the function

$$F(x)*G(x)=\int_{-\pi}^{\pi}F(x-y)G(y)dy$$

is called the convolution of F and G on the interval $(-\pi, \pi)$ (compare [12] p. 270 and [5] p. 274.)

Lemma 3. If $F(x)$ and $G(x)$ are sectionally continuous and $F(x+2\pi)=-F(x)$ and $G(x+2\pi)=-G(x)$, then

$$F(x)*G(x)=G(x)*F(x).$$

Theorem 6. If $F(x)$ and $G(x)$ are sectionally continuous and even functions and if $F(x+2\pi)=-F(x)$, then

$$C\{F(x)\}C\{G(x)\}=\tfrac{1}{2}C\{F(x)*G(x)\}.$$

Theorem 7. If $F(x)$ and $G(x)$ are sectionally continuous and odd functions and if $F(x+2\pi)=-F(x)$, then

$$S\{F(x)\}S\{G(x)\}=-\tfrac{1}{2}C\{F(x)*G(x)\}.$$

Theorem 8. If $F(x)$ is an even and $G(x)$ is an odd sectionally continuous function and if $F(x+2\pi)=-F(x)$, then

$$C\{F(x)\}S\{G(x)\}=\tfrac{1}{2}S\{F(x)*G(x)\}.$$

Theorem 9. If $F(x)$ is an odd and $G(x)$ is an even sectionally continuous function and if $F(x+2\pi) = -F(x)$, then

$$S\{F(x)\}C\{G(x)\} = \tfrac{1}{2}S\{F(x)*G(x)\}.$$

Remark. If in the above four theorems $G(x)$ satisfies the condition $G(x+2\pi) = -G(x)$, then according to lemma 3 the convolution is commutative. That being the case theorems 8 and 9 say the same thing.

Example. Given $q_c(n-\frac{1}{2}) = (n-\frac{1}{2})^{-2}[1-(n-\frac{1}{2})^2]^{-1}$, find $Q(x)$. $q_c(n-\frac{1}{2}) = S\{x\} \times S\{\sin x\}$ (see tables section 12). According to theorem 7 if $F(x) = x$ in $(0, \pi)$ and

$$F(x + 2\pi) = -F(x), \qquad G(x) = \sin x \quad \text{in} \quad (-\infty, \infty)$$

$$Q(x) = -\frac{1}{2}\left[\int_{-\pi}^{x-\pi} (2\pi - x + y)\sin y\,dy + \int_{x-\pi}^{\pi} (x - y)\sin y\,dy\right]$$

$$= \pi - x + \sin x.$$

8. Basic problem. In section 9 an application of the solution of the following problem will be made.

$$\frac{\partial Y}{\partial t} + \sum_{r=1}^{m} (-1)^r \frac{\partial^{(2r)} Y}{\partial x^{2r}} = 0, \qquad 0 < x < \pi, \qquad 0 < t,$$

$$\frac{\partial^{(2m-2s-2)} Y(0+, t)}{\partial x^{2m-2s-2}} = 0, \qquad \frac{\partial^{(2m-2s-1)} Y(\pi-, t)}{\partial x^{2m-2s-1}} = 0,$$

$$s = 0, 1, 2, \cdots, m-1, \qquad Y(x, 0+) = 1.$$

The S-transformation (19) with respect to x gives according to (23) the following transformed problem:

$$\frac{dy(n-\frac{1}{2}, t)}{dt} + y(n-\tfrac{1}{2}, t)\sum_{r=1}^{m} (n-\tfrac{1}{2})^{2r} = 0$$

$$y(n-\tfrac{1}{2}, 0+) = S\{1\} = 1/(n-\tfrac{1}{2}).$$

The solution of this transformed problem is

$$y(n-\tfrac{1}{2}, t) = (n-\tfrac{1}{2})^{-1} \exp\left[-\sum_{r=1}^{m} t(n-\tfrac{1}{2})^{2r}\right].$$

According to (21) the solution of the basic problem is thus

$$Y(x, t) = (2/\pi)\sum_{n=1}^{\infty} (n-\tfrac{1}{2})^{-1} \exp\left[-\sum_{r=1}^{m} t(n-\tfrac{1}{2})^{2r}\right] \sin (n-\tfrac{1}{2})x.$$

9. A generalized heat conduction problem. A formal solution of the following generalized heat conduction problem is obtained by use of the operator C when $K = \{n-\frac{1}{2}\}$ (see 20)).

$$L(U) = \frac{\partial U}{\partial t} + A_0(t)U(x, t) + \sum_{r=1}^{m} A_{2r}(t)(-1)^r \frac{\partial^{(2r)} U}{\partial x^{2r}} = Q(x, t),$$

$$0 < x < \pi, \qquad t > 0, \qquad A_{2r}(t) > 0,$$

$$\frac{\partial^{(2m-2s-1)} U(0+, t)}{\partial x^{2m-2s-1}} = B_{2m-2s-1}(t),$$

$$\frac{\partial^{(2m-2s-2)} U(\pi-, t)}{\partial x^{2m-2s-2}} = D_{2m-2s-2}(t), \qquad s = 0, 1, 2, \cdots, m-1,$$

$$U(x, 0+) = F(x).$$

This problem can be resolved into $2m+2$ problems each of which has $2m+1$ homogeneous conditions and one non-homogeneous condition. The sum of the solutions of these problems is the solution of the given problem. There are four essentially different types of problems, a formal solution of which is given in the following.

Problem I. $L(U) = Q(x, t)$, $0 < x < \pi$, $0 < t$, $0 < A_{2r}(t)$,

$$\frac{\partial^{(2m-2s-1)} U(0+, t)}{\partial x^{2m-2s-1}} = 0, \qquad \frac{\partial^{(2m-2s-2)} U(\pi-, t)}{\partial x^{2m-2s-2}} = 0,$$

$$s = 0, 1, 2, \cdots, m-1, \qquad U(x, 0+) = 0.$$

The C-transformation (20) applied with respect to x leads according to (30) to the following transformed problem:

$$\frac{du_c(n-\tfrac{1}{2}, t)}{dt} + \left[A_0(t) + \sum_{r=1}^{m} A_{2r}(t)(n-\tfrac{1}{2})^{2r}\right] u_c(n-\tfrac{1}{2}, t) - q_c(n-\tfrac{1}{2}, t) = 0,$$

$$u_c(n-\tfrac{1}{2}, 0+) = 0.$$

This equation is of the type $y'(x) + P(x)y(x) - Q(x) = 0$ with initial condition $y(0) = 0$. The solution thereof can be written in the form

$$y(x) = \int_0^x \exp\left[-p(x, w)\right] Q(w)dw, \quad \text{where} \quad p(x, w) = \int_w^x P(v)dv.$$

Hence the solution of the transformed problem can be written in the form

$$u_c(n-\tfrac{1}{2}, t) = \int_0^t \exp\left[-\overline{A}_0(t, v)\right] \exp\left[-\sum_{r=1}^{m} \overline{A}_{2r}(t, v)(n-\tfrac{1}{2})^{2r}\right] q_c(n-\tfrac{1}{2}, v)dv,$$

where

$$\overline{A}_{2r}(t, v) = \int_v^t A_{2r}(y)dy.$$

Now

$$\exp\left[-\sum_{r=1}^{m} \overline{A}_{2r}(t, v)(n-\tfrac{1}{2})^{2r}\right] = (n-\tfrac{1}{2})S\left\{Y(x, \overline{A}_{2r}(t, v))\right\},$$

where Y is the solution of the basic problem of section 8. And according to (15) with $k = n - \frac{1}{2}$ and $r = 1$.

$$(n - \tfrac{1}{2})S\{Y(x, \overline{A}_{2r}(t, v))\} = C\{Y_x(x, \overline{A}_{2r}(t, v))\},$$

so that

$$u_c(n - \tfrac{1}{2}, t) = \int_0^t \exp\,[-\overline{A}_0(t, v)]C\{Y_x(x, \overline{A}_{2r}(t, v))\}C\{Q(x, v)\}dv.$$

Using theorem 6 the solution of problem I can be written in the following form:

$$U(x, t) = \int_0^t \exp\,[-\overline{A}_0(t, v)][\tfrac{1}{2}Y_x(x, \overline{A}_{2r}(t, v)) * Q_0(x, v)]dv,$$

where

$$Q_0(x, v) = Q(x, v) \qquad 0 \leqq x \leqq \pi$$
$$Q_0(-x, v) = Q_0(x, v), \qquad -\pi \leqq x \leqq \pi.$$

Problem II. $L(U) = 0$, $0 < x < \pi$, $0 < t$, $A_{2r}(t) > 0$,

$$\frac{\partial^{(2m-2s-1)}U(0+, t)}{\partial x^{2m-2s-1}} = B_{2m-2s-1}(t) \quad \text{for} \quad s = i,$$
$$= 0, \qquad \text{for} \quad s \neq i,$$
$$\frac{\partial^{(2m-2s-2)}U(\pi-, t)}{\partial x^{2m-2s-2}} = 0, \qquad U(x, 0+) = 0,$$
$$s = 0, 1, 2, \cdots, m - 1.$$

According to (24) the C-transformation applied with respect to x leads to the following transformed problem.

$$\frac{du_c(n - \tfrac{1}{2}, t)}{dt} + \left[A_0(t) + \sum_{r=1}^{m} A_{2r}(t)(n - \tfrac{1}{2})^{2r}\right]u_c(n - \tfrac{1}{2}, t)$$
$$+ \sum_{r=0}^{m} A_{2r}(t)(-1)^r(-1)^{i+1}(n - \tfrac{1}{2})^{2i}B_{2r-2i-1}(t) = 0,$$
$$u_c(n - \tfrac{1}{2}, 0+) = 0.$$

With the notation of problem I the solution of this transformed problem reads:

$$u_c(n - \tfrac{1}{2}, t) = \int_0^t \exp\,[-\overline{A}_0(t, v)] \exp\left[-\sum_{r=1}^{m} \overline{A}_{2r}(t, v)(n - \tfrac{1}{2})^{2r}\right](n - \tfrac{1}{2})^{2i}(-1)^i$$
$$\cdot \sum_{r=0}^{m} (-1)^r B_{2r-2i-1}(v)A_{2r}(v)dv.$$

By use of the solution $Y(x, t)$ of the basic problem and (15) with $r = i+1$ and $k = n - \tfrac{1}{2}$

$$(n - \tfrac{1}{2})^{2i} \exp\left[-\sum_{r=1}^{m} \overline{A}_{2r}(t, v)(n - \tfrac{1}{2})^{2r}\right] = (n - \tfrac{1}{2})^{2i+1}S\{Y(x, \overline{A}_{2r}(t, v))\}$$
$$= (-1)^{i+2}C\left\{\frac{\partial^{(2i+1)}Y(x, \overline{A}_{2r}(t, v))}{\partial x^{2i+1}}\right\}.$$

Hence the solution of problem II reads

$$U(x,t)=\int_0^t \exp\left[-\overline{A}_0(t,v)\right]\frac{\partial^{(2i+1)}Y(x,\overline{A}_{2r}(t,v))}{x^{2i+1}}\left(\sum_{r=0}^{m}(-1)^r A_{2r}(v)B_{2r-2i-1}(v)\right)dv.$$

Problem III. $L(U)=0$, $0<x<\pi$, $0<t$, $A_{2r}(t)>0$,

$$\frac{\partial^{(2m-2s-1)}U(0+,t)}{\partial x^{2m-2s-1}}=0, \qquad s=0,1,2,\cdots,m-1,$$

$$\frac{\partial^{(2m-2s-2)}U(\pi-,t)}{\partial x^{2m-2s-2}}=D_{2m-2s-2}(t), \quad s=i,$$

$$=0, \qquad s\neq i,$$

$$U(x,0+)=0,$$

According to (24) the transformed problem reads:

$$\frac{du_c(n-\frac{1}{2},t)}{dt}+\left[A_0(t)+\sum_{r=1}^{m}A_{2r}(t)(n-\tfrac{1}{2})^{2r}\right]u_c(n-\tfrac{1}{2},t)$$

$$+\sum_{r=0}^{m}(-1)^r A_{2r}(t)(-1)^{i+1}(-1)^n(n-\tfrac{1}{2})^{2i+1}D_{2r-2i-2}(t)=0$$

$$u_c(n-\tfrac{1}{2},0+)=0.$$

The solution thereof in the notation of problem I is

$$u_c(n-\tfrac{1}{2},t)=\int_0^t \exp\left[-\overline{A}_0(t,v)\right]\exp\left[-\sum_{r=1}^{m}\overline{A}_{2r}(t,v)(n-\tfrac{1}{2})^{2r}\right](-1)^{n+i}(n-\tfrac{1}{2})^{2i+1}$$

$$\cdot\sum_{r=0}^{m}(-1)^r A_{2r}(v)D_{2r-2i-2}(v)dv.$$

By use of the solution $Y(x,t)$ of the basic problem and (23) with $r=i+1$

$$(n-\tfrac{1}{2})^{2i+1}\exp\left[-\sum_{r=1}^{m}\overline{A}_{2r}(t,v)(n-\tfrac{1}{2})^{2r}\right]=(n-\tfrac{1}{2})^{2i+2}S\left\{Y(x,\overline{A}_{2r}(t,v))\right\}$$

$$=(-1)^{i+1}S\left\{\frac{\partial^{(2i+2)}Y(x,\overline{A}_{2r}(t,v))}{\partial x^{2i+2}}\right\}.$$

Using theorem 3

$$(-1)^n S\left\{\frac{\partial^{(2i+2)}Y(x,\overline{A}_{2r}(t,v))}{\partial x^{2i+2}}\right\}=-C\left\{\frac{\partial^{(2i+2)}Y(y,\overline{A}_{2r}(t,v))}{y^{2i+2}}\right\},$$

$$y=\pi-x$$

so that the solution of problem III reads

$$U(x, t) = \int_0^t \exp\left[-\overline{A}_0(t, v)\right] \frac{\partial^{(2i+2)} Y(y, \overline{A}_{2r}(t, v))}{\partial y^{2i+2}} \left[\sum_{r=0}^{m} (-1)^r A_{2r}(v) D_{2r-2i-2}(v)\right] dv,$$

$$y = \pi - x.$$

Problem IV. $L(U) = 0$, $0 < x < \pi$, $0 < t$, $A_{2r}(t) > 0$,

$$\frac{\partial^{(2m-2s-1)} U(0+, t)}{\partial x^{2m-2s-1}} = 0, \qquad \frac{\partial^{(2m-2s-2)} U(\pi-, t)}{\partial x^{2m-2s-2}} = 0,$$

$$s = 0, 1, 2, \cdots, m-1, \qquad U(x, 0+) = F(x).$$

According to (24) the transformed problem reads

$$\frac{du_c(n-\frac{1}{2}, t)}{dt} + \left[A_0(t) + \sum_{r=1}^{m} A_{2r}(t)(n-\tfrac{1}{2})^{2r}\right] u_c(n-\tfrac{1}{2}, t) = 0,$$

$$u_c(n-\tfrac{1}{2}, 0+) = f_c(n-\tfrac{1}{2}).$$

The solution thereof in the notation of problem I can be written as

$$u_c(n-\tfrac{1}{2}, t) = f_c(n-\tfrac{1}{2}) \exp\left[-\overline{A}_0(t, 0)\right] \exp\left[-\sum_{r=1}^{m} \overline{A}_{2r}(t, 0)(n-\tfrac{1}{2})^{2r}\right]$$

$$= \exp\left[-\overline{A}_0(t, 0)\right] C\{F(x)\} C\left\{\frac{\partial Y(x, \overline{A}_{2r}(t, 0))}{\partial x}\right\},$$

since

$$\exp\left[-\sum_{r=1}^{m} \overline{A}_{2r}(t, 0)(n-\tfrac{1}{2})^{2r}\right] = (n-\tfrac{1}{2}) S\{Y(x, \overline{A}_{2r}(t, 0))\}$$

$$= C\left\{\frac{\partial Y(x, \overline{A}_{2r}(t, 0))}{\partial x}\right\}$$

according to (15) with $r=1$ and $k=n-\frac{1}{2}$.

By theorem 6 the solution of problem IV can be written as

$$U(x, t) = \exp\left[-\overline{A}_0(t, 0)\right]\left[\frac{1}{2} \frac{\partial Y(x, \overline{A}_{2r}(t, 0))}{\partial x} * F_0(x)\right],$$

where $F_0(x) = F(x)$ in $(0, \pi)$ and $F_0(-x) = F_0(x)$ in $(-\pi, \pi)$.

It can be seen now that the solution of each problem is expressed in closed form in terms of the solution of the basic problem of section 8. Thus *the solution of the given problem* being the sum of the 2m+2 problems *is expressed in closed form* in terms of the solution of the basic problem.

10. A problem in heat conduction. As a particular case ($m=1$) of section 9 the solution of the following problem can be obtained.

To find the temperature distribution $U(x, t)$ in a slab of length π with a heat source inside, The end $x=\pi$ is kept at a temperature $D(t)$ and radiation through the end $x=0$ at a rate $B(t)$ takes place. Furthermore there is radiation through the lateral surface and the thermal diffusivity $K(t)$ depends on time. The initial temperature of the slab is $F(x)$.(Compare also [1].)

The mathematical formulation of this problem is the problem of section 9 with $m=1$, $A_2(t)=K(t)$, $A_0(t)=A(t)$, $B_1(t)=B(t)$, $D_0(t)=D(t)$, i.e.

$$\frac{\partial U}{\partial t} - K(t)\frac{\partial^2 U}{\partial x^2}A(t)U(x,t) = Q(x,t), \qquad 0<x<\pi, \qquad 0<t,$$

$$\frac{\partial U(0+,t)}{\partial x} = B(t) \qquad U(\pi-,t) = D(t),$$

$$U(x,0+) = F(x).$$

Put

$$\overline{K}(t,v) = \int_v^t K(x)dx, \qquad \overline{A}(t,v) = \int_v^t A(x)dx.$$

The solution of the basic problem of section 8 with $m=1$ is

$$Y(x,t) = (2/\pi)\sum_1^\infty (n-\tfrac{1}{2})^{-1}\exp\left[-(n-\tfrac{1}{2})^2 t\right]\sin(n-\tfrac{1}{2})x.$$

The solution of the heat conduction problem can be written in closed form using Jacobi's ϑ_2-function.
For

$$\frac{\partial Y(x,t)}{\partial x} = (2/\pi)\sum_1^\infty \exp\left[-(n-\tfrac{1}{2})^2 t\right]\cos(n-\tfrac{1}{2})x$$

$$= (1/\pi)\vartheta_2(x/2, \exp[-t]), \quad \text{(see [12] p. 464.)}$$

Thus the solution reads

$$U(x,t) = \int_0^t \exp\left[-\overline{A}(t,v)\right]\left[(1/2\pi)\vartheta_2(x/2, \exp\left\{-\overline{K}(t,v)\right\}) * Q_0(x,v)\right]dv$$

$$- (1/\pi)\int_0^t \exp\left[-\overline{A}(t,v)\right]\vartheta_2(x/2, \exp\left\{-\overline{K}(t,v)\right\})K(v)B(v)dv$$

$$- (1/\pi)\int_0^t \exp\left[-\overline{A}(t,v)\right]\frac{\partial \vartheta_2}{\partial y}(y/2, \exp\left\{-\overline{K}(t,v)\right\})K(v)D(v)dv$$

$$y = \pi - x$$

$$+ \exp\left[-\overline{A}(t,0)\right]\left[(1/2\pi)\vartheta_2(x/2, \exp\left\{-\overline{K}(t,0)\right\}) * F_0(x)\right].$$

Remarks. Since the ϑ-functions are tabulated the above form of the solution can be used to determine by mechanical integration the numerical values of $U(x,t)$ for given values x and t.

For questions on uniqueness and existence of solutions in heat conduction problems the reader is referred to [9].

11. An illustration of the method in the case of double characteristic values. Summary of the operational method. A formal solution of the following boundary value problem is to be found.

$$L(U) = U_t(x, t) - K(t)U_{xx}(x, t) + A(t)U(x, t) = Q(x, t),$$
$$0 < x < \pi, \qquad 0 < t, \qquad A(t) > 0, \qquad K(t) > 0,$$
$$L_1(U) = U_x(0+, t) + U_x(\pi -, t) = B(t),$$
$$L_2(U) = U(0+, t) + U(\pi -, t) = D(t),$$
$$U(x, 0+) = F(x).$$

As kernel of the transformation (9) to be applied to this problem the characteristic functions of $y''(x)+k^2y(x)=0$, $y'(0)+y'(\pi)=0$, $y(0)+y(\pi)=0$, are used. In this case the determinant $D(k)$ is of rank zero. The double characteristic values are $2n-1$, $n=1, 2, \cdots$, and $\sin(2n-1)x$ as well as $\cos(2n-1)x$ are characteristic functions. This means the sine as well as the cosine coefficients of the solution of the above problem have to be found. The formulae to be used are:

$$S\{F''(x)\} = -(2n-1)^2 f_s(2n-1) + (2n-1)[F(0)+F(\pi)], \tag{25}$$
$$C\{F''(x)\} = -(2n-1)^2 f_c(2n-1) - [F'(0)+F'(\pi)], \tag{26}$$
$$C\{F'(x)\} = (2n-1) f_s(2n-1) - [F(\pi)+F(0)]. \tag{27}$$

The last three equations follow from (12), (13) and (15) respectively with $r=1$ and $k=2n-1$, $n=1, 2, \cdots$. Relation (11) yields

$$S^{-1}\{f_s(2n-1)\} = (2/\pi)\sum_1^\infty f_s(2n-1)\sin(2n-1)x \quad \text{in} \quad (0, \pi). \tag{28}$$

Furthermore the solution of the following auxiliary problem will be useful.

$$Y_t(x, t) - Y_{xx}(x, t) = 0, \quad \text{in} \quad (0, \pi), \qquad Y_x(0+, t) + Y_x(\pi -, t) = 0,$$

$Y(0+, t)+Y(\pi-, t)=0$, $Y(x, 0+)=1$. From (25) it follows that $[dy_s(2n-1, t)/dt]+(2n-1)^2y_s(2n-1, t)=0$, $y_s(2n-1, 0+)=S\{1\}=2/(2n-1)$. Hence $y_s(2n-1, t)=2\exp[-t(2n-1)^2](2n-1)^{-1}$, which gives the sine coefficient of the solution of the auxiliary problem. The cosine coefficient is zero since $C\{1\}=0$ when $k=2n-1$; so that the solution of the auxiliary problem is according to (28)

$$Y(x, t) = (2/\pi)\sum_1^\infty 2(2n-1)^{-1}\exp[-(2n-1)^2 t]\sin(2n-1)x, \tag{29}$$

$$Y_x(x, t) = (2/\pi)\vartheta_2(x, \exp[-4t]), \quad \text{see [18] p. 464.} \tag{30}$$

The following formulae by Kniess (see [12]) will also be used. If $K=\{n\}$ then a) if $F(x)$ is even and periodic with period 2π and $G(x)$ is odd, then

$$C\{F(x)\}S\{G(x)\} = \tfrac{1}{2}S\{F(x)*G(x)\}; \tag{31}$$

b) if $F(x)$ and $G(x)$ are even and $F(x)$ is periodic with period 2π, then

$$C\{F(x)\}C\{G(x)\} = \tfrac{1}{2}C\{F(x)*G(x)\}. \tag{32}$$

The solution of the given problem can be written as the sum of the solutions of the following four problems, where $0<x<\pi$, $0<t$, $A(t)>0$, $K(t)>0$.

$$\text{I. } L(U) = Q(x, t), \qquad L_1(U) = L_2(U), \qquad U(x, 0+) = 0.$$
$$\text{II. } L(U) = L_2(U), \quad U(x, 0+) = 0, \qquad L_1(U) = B(t).$$
$$\text{III. } L(U) = L_1(U), \quad U(x, 0+) = 0, \qquad L_2(U) = D(t).$$
$$\text{IV. } L(U) = L_1(U) = L_2(U) = 0, \qquad U(x, 0+) = F(x).$$

The sine and the cosine transforms using (25) and (26) respectively of the solution of each of these problems have to be found. The sine transform of the solution of II and the cosine transform of the solution of III are zero.

With the notation of section 10 it follows from (25) that the sine transform of the solution of I is

$$u_s(2n-1, t) = \int_0^t \exp\left[-\overline{A}(t, v)\right] \exp\left[-(2n-1)^2\overline{K}(t, v)\right] q_s(2n-1, v)dv.$$

Relations (29) and (27) yield

$$\exp\left[-(2n-1)^2 t\right] = \left[(2n-1)/2\right] S\left\{Y(x, t)\right\} = \tfrac{1}{2} C\left\{Y_x(x, t)\right\}.$$

Using (31) and (30) the sine part of the solution of I can be written as

$$(1/2\pi)\int_0^t \exp\left[-\overline{A}(t, v)\right]\left[\vartheta_2(x, \exp\left(-4\overline{K}(t, v)\right\}) * Q_0(x\ v)\right]dv,$$

where Q is extended so as to be odd with respect to x.

From (26), (29) and (27) it follows that

$$u_c(2n-\tfrac{1}{2}, t) = \frac{1}{2}\int_0^t \exp\left[-\overline{A}(t, v)\right] C\left\{Y_x(x\ \overline{K}(t, v))\right\} C\left\{Q(x\ v)\right\} dv.$$

According to (32) and (30) the cosine part of the solution of I can be written as

$$(1/2\pi)\int_0^t \exp\left[-\overline{A}(t, v)\right]\left[\vartheta_2(x, \exp\left\{-4\overline{K}(t, v)\right\}) * Q_e(x, v)\right]dv,$$

where Q is extended so as to be even with respect to x.

It may be left to the reader to show that in a similar way the sine and cosine parts of the problems II–IV can be expressed in closed form in terms of ϑ_2 by using relations (25) to (32). Thus the solution of the given problem can be expressed in closed form in terms of Jacobi's ϑ_2 function as follows:

$$\begin{aligned}
U(x, t) = {} & (1/2\pi)\int_0^t \exp\left[-\overline{A}(t, v)\right]\left[\vartheta_2(x, \exp\left\{-4\overline{K}(t, v)\right\}) * Q_0(x, v)\right]dv \\
& + (1/2\pi)\int_0^t \exp\left[-\overline{A}(t, v)\right]\left[\vartheta_2(x, \exp\left\{-4\overline{K}(t, v)\right\}) * Q_e(x, v)\right]dv \\
& - (1/\pi)\int_0^t \exp\left[-\overline{A}(t, v)\right]\vartheta_2(x, \exp\left\{-4\overline{K}(t, v)\right\}) B(v)K(v)dv \\
& - (1/\pi)\int_0^t \exp\left[-\overline{A}(t, v)\right]\vartheta_2(x, \exp\left\{-4\overline{K}(t, v)\right\}) D(v)K(v)dv \\
& + (1/2\pi)\exp\left[-\overline{A}(t, 0)\right]\left[\vartheta_2(x, \exp\left\{-4\overline{K}(t, 0)\right\}) * F_0(x)\right] \\
& + (1/2\pi)\exp\left[-\overline{A}(t, 0)\right]\left[\vartheta_2(x, \exp\left\{-4\overline{K}(t, 0)\right\}) * F_e(x)\right].
\end{aligned}$$

Summary of the operational method. The method is applicable to the following type of boundary value problem:

$$\sum_{r=0}^{m} (-1)^r A_{2r} \frac{d^{2r}F(x)}{dx^{2r}} = G(x) \quad \text{in} \quad (0, \pi),$$

where $G(x)$, the quantities

$$a_1F^{(2r-2s-2)}(0) + a_2F^{(2r-2s-1)}(0) + a_3F^{(2r-2s-2)}(\pi) + a_4F^{(2r-2s-2)}(\pi)$$

and

$$b_1F^{(2r-2s-2)}(0) + b_2F^{(2r-2s-1)}(0) + b_3F^{(2r-2s-2)}(\pi) + b_4F^{(2r-2s-2)}(\pi)$$

and the constants A_{2r} are assigned, $s=0, 1, 2, \cdots, r-1$. The constants a_i, b_i, $i=1, 2, 3, 4$ are allowed to be different for each s. The transformation is also useful in boundary value problems in partial differential equations, where one or more variables behave like x in the above problem. In order to obtain a formal solution quickly, and possibly a closed form solution, the following procedure is recommended:

1° Set up (3) corresponding to the given problem.

2° Find the roots of $D(k)=0$ (see (7)).

3° Find the rank of the determinant $D(k)$.

4° If the rank is greater than zero set up $T\{F\}$ using (9). If the rank is zero set up $T_1\{F\}$ and $T_2\{F\}$, where T_1 and T_2 have the two independent characteristic functions of (3) as respective kernels.

5° Find $T\{F^{2r}\}$ using (16).

6° Apply $T\{F^{2r}\}$ to the boundary value problem (which may be resolved into several problems each of which has only one non-homogeneous condition) and find $f(k_n)$, the transform of the solution.

7° Try to obtain the inverse $T^{-1}\{f(k_n)\}$ in closed form by application of the theorems of sections 3, 5, 7 or by use of tables of transforms or a combination of both.

8° If 7° does not lead to the solution, use (11) to find $T^{-1}\{f(k_n)\}$ and obtain the solution in series form.

Remark. Formulae for a few special cases of T can be found in the tables section 12.

12. Tables. Tables A–D contain a few examples of transforms. It would be desirable to have extensive tables of transforms, since they would help in obtaining closed form solutions of boundary value problems (see the example of section 6).

Table E below contains a list of transformations which are special cases of the transformation T (see (9)) of section 1. In each case the formula for the transform of an even derivative (see (16)) as well as the inverse transformation (see (11)) are given. The completion of this list of special transformations is left to the reader.

A. Tables of S-transforms.

	$f_s(k)=S\{F(x)\}$	$F(x)$
1.	$(1-\cos k\pi)k^{-1}$	1
2.	$k^{-2}\sin k\pi-k^{-1}\pi\cos k\pi$	x
3.	$[(-1)^r(2r)!(1-\cos k\pi)/k^{2r+1}]+\cos k\pi\sum_{s=0}^{r-1}(-1)^{r+s}[(2r)!\pi^{2s+2}/(2s+2)!k^{2r-2s-1}]$ $+\sin k\pi\sum_{s=0}^{r-1}(-1)^{r+s+1}[(2r)!\pi^{2s+1}/(2s+1)!k^{2r-2s}]$	x^{2r}
4.	$\sin k\pi\sum_{s=0}^{r-1}(-1)^{r+s+1}[\pi^{2s}(2r-1)!/(2s)!k^{2r-2s}]$ $+\cos k\pi\sum_{s=0}^{r-1}(-1)^{s+r}[\pi^{2r+1}(2r-1)!/(2s+1)!k^{2r-2s-1}]$	x^{2r-1}
5.	$(k^2-a^2)^{-1}(a\cos a\pi\sin k\pi-k\sin a\pi\cos k\pi)$, $a\neq k$	$\sin ax$
6.	$(k^2-a^2)^{-1}(k-k\cos k\pi\cos a\pi-a\sin k\pi\sin a\pi)$, $a\neq k$	$\cos ax$

B. Tables of C-transforms.

	$f_c(k)=C\{F(x)\}$	$F(x)$
1.	$k^{-1}\sin k\pi$	1
2.	$k^{-1}\pi\sin k\pi-(1-\cos k\pi)k^{-2}$	x
3.	$\sin k\pi\sum_{s=0}^{r}(-1)^{r+s}[\pi^{2s}(2r)!/(2s)!k^{2r-2s+1}]$ $+\cos k\pi\sum_{s=0}^{r-1}(-1)^{r+s+1}[\pi^{2s+1}(2r)!/(2s+1)!k^{2r-2s}]$	x^{2r}
4.	$[(-1)^r(2r-1)!/k^{2r}]+\cos k\pi\sum_{s=0}^{r-1}(-1)^{r+s+1}[(2r-1)!\pi^{2s}/(2s-2)!k^{2s-2r}]$ $+\sin k\pi\sum_{s=0}^{r-1}(-1)^{r+s+1}[(2r-1)!\pi^{2s+1}/(2s-1)!k^{2r-2s-1}]$	x^{2r-1}
5.	$(k^2-a^2)^{-1}(-a+k\sin k\pi\sin a\pi+a\cos k\pi\cos a\pi)$, $a\neq k$	$\sin ax$
6.	$(k^2-a^2)^{-1}(k\cos a\pi\sin k\pi-a\sin a\pi\cos k\pi)$, $a\neq k$	$\cos ax$

C. Table of S-transforms when $K=\{n-\frac{1}{2}\}$, $n=1, 2, \cdots$.

	$f_s(n-\frac{1}{2})=S\{F(x)\}$	$F(x)$
1.	$(n-\frac{1}{2})^{-1}$	1
2.	$(-1)^{n+1}(n-\frac{1}{2})^{-2}$	x
3.	$(n-\frac{1}{2})^{-3}$	$x(2\pi-x)/2$
4.	$(n-\frac{1}{2})^{-1}(1-(n-\frac{1}{2})^2)^{-1}$	$1-\cos x$
5.	$(n-\frac{1}{2})[(n-\frac{1}{2})^2-a^2]^{-1}$	$\cos a(\pi-x)/\cos a\pi$
6.	$(-1)^{n+1}a[a^2+(n-\frac{1}{2})^2]^{-1}$	$\operatorname{sh} ax/\cosh a\pi$

D. Tables of C-transforms when $K=\{n-\frac{1}{2}\}$, $n=1, 2, \cdots$.

	$f_c(n-\frac{1}{2})=C\{F(x)\}$	$F(x)$
1.	$(-1)^{n+1}(n-\frac{1}{2})^{-1}$	1
2.	$(-1)^{n+1}\pi(n-\frac{1}{2})^{-1}-(n-\frac{1}{2})^{-2}$	x
3.	$(n-\frac{1}{2})^{-2}$	$\pi-x$
4.	$(n-\frac{1}{2})^{-2}[1-(n-\frac{1}{2})^2]^{-1}$	$\pi-x-\sin x$
5.	$[(n-\frac{1}{2})^2-a^2]^{-1}$	$\sin a(\pi-x)/\cos a\pi$
6.	$(-1)^{n+1}(n-\frac{1}{2})[(n-\frac{1}{2})^2-a^2]^{-1}$	$\cos ax/\cos a\pi$
7.	$[a^2+(n-\frac{1}{2})^2]^{-1}$	$\operatorname{sh} a(\pi-x)/\cosh a\pi$
8.	$(-1)^{n+1}(n-\frac{1}{2})[a^2+(n-\frac{1}{2})^2]^{-1}$	$\cosh ax/\cosh a\pi$

E. Table of some special transformations.

$$S\{F(x)\} = \int_0^\pi F(x) \sin nx dx = f_s(n), \qquad n = 1, 2, \cdots \tag{33}$$

$$S^{-1}\{f_s(n)\} = (2/\pi)\sum_1^\infty f_s(n) \sin nx \quad \text{in} \quad (0, \pi)$$

$$S\{F^{(2r)}(x)\} = (-n^2)^r f_s(n) + \sum_{s=0}^{s=r-1} (-1)^s n^{2s+1}[F^{(2r-2s-2)}(0) - (-1)^n F^{(2r-2s-2)}(\pi)]$$

$$C\{F(x)\} = \int_0^\pi F(x) \cos nx dx = f_c(n), \qquad n = 0, 1, 2, \cdots \tag{34}$$

$$C^{-1}\{f_c(n)\} = (1/\pi)f_c(0) + (2/\pi)\sum_1^\infty f_c(n) \cos nx \quad \text{in} \quad (0, \pi)$$

$$C\{F^{(2r)}(x)\} = (-n^2)^r f_c(n) + \sum_{s=0}^{s=r-1} (-1)^s n^{2s}[(-1)^n F^{(2r-2s-1)}(\pi) - F^{(2r-2s-1)}(0)]$$

$$S\{F(x)\} = \int^\pi F(x) \sin (n - \tfrac{1}{2})x dx = f_s(n - \tfrac{1}{2}), \qquad n = 1, 2, \cdots \tag{35}$$

$$S^{-1}\{f_s(n - \tfrac{1}{2})\} = (2/\pi)\sum_1^\infty f_s(n - \tfrac{1}{2}) \sin (n - \tfrac{1}{2})x \quad \text{in} \quad (0, \pi)$$

$$S\{F^{(2r)}(x)\} = (-(n - \tfrac{1}{2})^2)^r f_s(n - \tfrac{1}{2}) + \sum_{s=0}^{s=r-1} (-1)^s (n - \tfrac{1}{2})^{2s}[(-1)^{n+1} F^{(2r-2s-1)}(\pi) + (n - \tfrac{1}{2})F^{(2r-2s-2)}(0)]$$

$$C\{F(x)\} = \int_0^\pi F(x) \cos (n - \tfrac{1}{2})x dx = f_c(n - \tfrac{1}{2}), \quad n = 1, 2, \cdots \tag{36}$$

$$C^{-1}\{f_c(n - \tfrac{1}{2})\} = (2/\pi)\sum_1^\infty f_c(n - \tfrac{1}{2}) \cos (n - \tfrac{1}{2})x \quad \text{in} \quad (0, \pi)$$

$$C\{F^{(2r)}(x)\} = (-(n - \tfrac{1}{2})^2)^r f_c(n - \tfrac{1}{2}) + \sum_{s=0}^{s=r-1} (-1)^{s+1}(n - \tfrac{1}{2})^{2s}[F^{(2r-2s-1)}(0) + (n - \tfrac{1}{2})(-1)^n F^{(2r-2s-2)}(\pi)]$$

$$S\{F(x)\} = \int_0^\pi F(x) \sin k_n x dx = f_s(k_n), \qquad n = 1, 2, \cdots, \tag{37}$$

where k_n are the roots of $\tan \pi = -(k/h)$, $k > 0$, $h \neq 0$

$$S^{-1}\{f_s(k_n)\} = \sum_1^\infty N(k_n) f_s(k_n) \sin k_n x \quad \text{in} \quad (0, \pi) \quad \text{and}$$

$$1/N(k_n) = (\pi/2) - (4k_n)^{-1} \sin 2\pi k_n$$

$$S\{F^{(2r)}(x)\} = (-k_n^2)^r f_s(k_n) + \sum_{s=0}^{s=r-1} (-k_n^2)^s [\sin k_n\pi [F^{(2r-2s-1)}(\pi) + hF^{(2r-2s-2)}(\pi)] + k_n F^{(2r-2s-2)}(0)]$$

$$\tilde{S}\{F(x)\} = \int_0^\pi F(x) \sin k_n(\pi - x)dx = \tilde{f}_s(k_n), \qquad n = 1, 2, \cdots \tag{38}$$

$$\tilde{S}^{-1}\{{}_s(k_n)\} = \sum_1^\infty N(k_n)\tilde{f}_s(k_n) \sin k_n(\pi - x), \quad \text{in} \quad (0, \pi),$$

k_n and $N(k_n)$ as in (37)

$$\tilde{S}\{F^{(2r)}(x)\} = (-k_n^2)^r f_s(k_n) + \sum_{s=0}^{s=r-1} (-1)^{s+1} k_n^{2s} [\sin k_n\pi [F^{(2r-2s-1)}(0) - hF^{(2r-2s-2)}(0)] - k_n F^{(2r-2s-2)}(\pi)]$$

$$C\{F(x)\} = \int_0^\pi F(x) \cos k_n x dx = f_c(k_n), \qquad n = 1, 2, \cdots, \tag{39}$$

where k_n are the roots of tank $\pi = (h/k)$, $k > 0$, $h \neq 0$

$$C^{-1}\{f_c(k_n)\} = \sum_1^\infty N(k_n) f_c(k_n) \cos k_n x \quad \text{in} \quad (0, \pi) \quad \text{and}$$

$$1/N(k_n) = (\pi/2) + (4k_n)^{-1} \sin 2k_n\pi$$

$$C\{F^{(2r)}(x)\} = (-k_n^2)^r f_c(k_n) + \sum_{s=0}^{s=r-1} (-k_n^2)^s [\cos k_n\pi [F^{(2r-2s-1)}(\pi) + hF^{(2r-2s-2)}(\pi)] - F^{(2r-2s-1)}(0)]$$

$$\tilde{C}\{F(x)\} = \int_0^\pi F(x) \cos k_n(\pi - x)dx = \tilde{f}_c(k_n), \qquad n = 1, 2, \cdots \tag{40}$$

$$\tilde{C}^{-1}\{f_c(k_n)\} = \sum_1^\infty N(k_n)\tilde{f}_c(k_n) \cos k_n(\pi - x), \quad \text{in} \quad (0, \pi)$$

k_n and $N(k_n)$ as in (39)

$$\tilde{C}\{F^{(2r)}(x)\} = (-k_n^2)^r \tilde{f}_c(k_n) + \sum_{s=0}^{s=r-1} (-1)^{s+1} k_n^{2s} [\cos k_n\pi [F^{(2r-2s-1)}(0) - hF^{(2r-2s-2)}(0)] - F^{(2r-2s-1)}(\pi)]$$

$$T\{F(x)\} = \int_0^\pi F(x)(a \sin k_n x - k_n b \cos k_n x)dx = f(k_n), \qquad n = 1, 2, \cdots \tag{41}$$

where k_n are the roots of tank $\pi = k(bc - ad)/(ac + k^2bd)$,

$$k > 0, \qquad ad - bc \neq 0$$

$$T^{-1}\{f(k_n)\} = \sum_1^\infty N(k_n)f(k_n)(a \sin k_n x - k_n b \cos k_n x) \quad \text{in} \quad (0, \pi) \quad \text{and}$$

$$1/N(k_n) = \tfrac{1}{2}(\pi a^2 + \pi k_n^2 b^2 - ab) + [(k_n^2 b^2 - a^2)/4k_n] \sin 2\pi k_n + (ab/2) \cos 2\pi k_n$$

$$T\{F^{(2r)}(x)\} = (-k_n^2)^r f(k_n) + \sum_{s=0}^{s=r-1} (-k_n^2)^s \{k_n[aF^{(2r-2s-2)}(0) + bF^{(2r-2s-1}(0)]$$

$$- k_n^2 b(c \sin k_n\pi + k_n d \cos k_n\pi)^{-1}[cF^{(2r-2s-2)}(\pi) + dF^{(2r-2s-1)}(\pi)]\}.$$

Bibliography

1. H. K. Brown, *Resolution of temperature problems by use of finite Fourier transformations*, Bull. Amer. Math. Soc. **50**, 376–385 (1944).

2. H. K. Brown, *The resolution of boundary value problems by means of the finite Fourier transformation: general vibration of a string*, J. Appl. Phys. **14**, 609–618 (1943).

3. H. K. Brown, *General vibration of a hinged beam*, J. Appl. Phys. **15**, 410–413 (1944).

4. R. V. Churchill, *Fourier series and boundary value problems*, McGraw-Hill, New York, 1941.

5. R. V. Churchill, *Modern operational mathematics in engineering*, McGraw-Hill, New York, 1944.

6. Courant-Hilbert, *Methoden der mathematischen Physik*, Berlin, 1931.

7. S. R. de Groot, *Sur l'intégration de quelques problèmes aux limites régis par l'équation de Fourier dite "de la chaleur" au moyen de la méthode des transformations fonctionnelles simultanées*, Nederl. Akad. Wetensch., Proc. **45**, 643–649, 820–825 (1942).

8. G. Doetsch, *Integration von Differentialgleichungen vermittels der endlichen Fourier Transformation*, Math. Ann. **112**, 52–68 (1935).

9. G. Doetsch, *Theorie und Anwendung der Laplace Transformation*, Dover Publications, New York, 1943.

10. Frank-v. Mises, *Differentialgleichungen der Physik*, 2nd ed. New York, 1943.

11. S. Kaczmarz and H. Steinhaus, *Theorie der Orthogonalreihen*, Warsaw, 1935.

12. H. Kniess, *Lösung von Randwertproblemen bei Systemen gewöhnlicher Differentialgleichungen vermittels der endlichen Fourier Transformation*, Math. Zeit. **44**, 266–291 (1938).

13. L. Koschmieder, *Die endliche Fouriersche Abbildung und ihr Nutzen bei Aufgaben der Wärmeleitung in Stäben und Platten*, Deutsche Math. **5**, 521–545 (1945).

14. M. Picone, *Appunti di analisi superiore*, Naples, 1940. pp. 790–793.

15. I. Roettinger, *An operational approach to the solution of boundary value problems by generalized Fourier series*, Abstract, Bull. Amer. Math. Soc. **51**, 67 (1945).

16. A. G. Strandhagen, *Use of sine transform for non-simply supported beams*, Quart. Appl. Math. **1**, 346–348 (1944).

17. C. J. Tranter, *On a problem in heat conduction*, Philos. Mag. (7) **35**, 102–105 (1944).

18. E. T. Whittaker and G. N. Watson, *A course of modern analysis*, Cambridge University Press, 1940.

19. A. Zygmund, *Trigonometrical series*, Warsaw, 1935.

NUMERICAL METHODS FOR FINDING CHARACTERISTIC ROOTS AND VECTORS OF MATRICES*

BY

W. M. KINCAID

University of Michigan

The present paper treats the problem of finding the characteristic roots and vectors of a matrix (having linear elementary divisors). The emphasis throughout is on methods of getting numerical results in practical cases rather than on theoretical questions.

Symmetric matrices are taken up first, and methods are discussed for finding (a) the largest characteristic root and corresponding vector, and (b) the other roots and vectors. All these methods are variants of the iteration process. They are then extended to general matrices, with particular reference to the case of complex roots. The paper closes with a brief discussion of the solution of algebraic equations by means of matrices.

Among earlier work on this subject, we may mention that of Hotelling[1] and Aitken.[2] Indeed, much of the material in the present paper is taken from Aitken's, through some modifications have been introduced, particularly with regard to the determination of roots other than the largest.

A recent paper by Fry[3] takes up matrices in connection with the solution of algebraic equations. The present paper, particularly the last section, is thus in a measure supplementary to Fry's. Wayland,[4] on the other hand, gives methods for reducing the problem of finding the roots of a given matrix to that of solving an algebraic equation; this matter is touched upon Sec. II.

Recent work along these lines has also been done by Morris[5] and Head.[6]

At this point the author wishes to acknowledge his indebtedness to Prof. W. Feller for many helpful suggestions made during the preparation of the manuscript.

I. Basic Relations and Definitions

1. Nature of the problem: definitions and notations. The problem of finding the characteristic roots and vectors of a matrix arises naturally in the solution of a system of linear equations of the type

$$\left.\begin{aligned} a_{11}x^{(1)} + a_{12}x^{(2)} + \cdots + a_{1n}x^{(n)} &= \lambda x^{(1)} \\ a_{21}x^{(1)} + a_{22}x^{(2)} + \cdots + a_{2n}x^{(n)} &= \lambda x^{(2)} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ a_{n1}x^{(1)} + a_{n2}x^{(2)} + \cdots + a_{nn}x^{(n)} &= \lambda x^{(n)} \end{aligned}\right\}. \tag{1}$$

* Received Dec. 27, 1946. This paper is a condensed version of a thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Brown University, October, 1946.

[1] H. Hotelling, Psychometrika **1**, 27–35 (1936).

[2] A. C. Aitken, Proc. Roy. Soc. Edinburgh **47**, 269–304 (1937).

[3] T. C. Fry, Q. Appl. Math. **3**, 89–105 (1945).

[4] H. Wayland, Q. Appl. Math. **2**, 277–306 (1944).

[5] J. Morris, Aircraft Engrg. **14**, 108–110 (1942).

[6] J. Morris and J. W. Head, Aircraft Engrg. **14**, 312–314, (1942).

In general, the system (1) has non-trivial solutions for n values $\lambda_1, \lambda_2, \cdots, \lambda_n$ of the parameter λ, which are called the *characteristic values* or *roots* of the system; these are the roots of the determinantal equation

$$\begin{vmatrix} a_{11}-\lambda & a_{21} & \cdots & a_{1n} \\ a_{12} & a_{22}-\lambda & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn}-\lambda \end{vmatrix} = 0. \tag{2}$$

Corresponding to any characteristic value λ_i there exists a solution $(x_i^{(1)}, x_i^{(2)}, \cdots, x_i^{(n)})$ which may be termed a *characteristic vector* of the system.

The vector $(x_i^{(1)}, \cdots, x_i^{(n)})$ may be regarded as an n by 1 matrix. Such a matrix is called a *column vector*. Similarly, a matrix having but one row is called a *row vector* and will be written in the form $\{y^{(1)}, y^{(2)}, \cdots, y^{(n)}\}$. In general, if V is any given column vector, we shall denote by V' the row vector having the same elements.

If $C=[c_{ij}]$ is an m by n matrix and $D=[d_{ij}]$ is an n by p matrix, their *product CD* is defined to be the m by p matrix E having elements of the form

$$e_{ij} = \sum_{k=1}^{n} c_{ik}d_{kj} \qquad (i = 1, 2, \cdots, m; j = 1, \cdots, p).$$

Note that DC is not necessarily equal to CD. In particular, a row vector R and a column vector K (with the same number of elements) have two products, a *scalar product* RK and a *matrix product* KR.

The system (1) may now be written in the form

$$AX = \lambda X, \tag{3}$$

where

$$A = \begin{bmatrix} a_{11} & a_{12} \cdots a_{1n} \\ a_{21} & a_{22} \cdots a_{2n} \\ \cdots & \cdots \\ a_{n2} & a_{n2} \cdots a_{nn} \end{bmatrix} \qquad X = \begin{bmatrix} x^{(1)} \\ x^{(2)} \\ \cdots \\ x^{(n)} \end{bmatrix}. \tag{4}$$

By a natural extension of meaning, the number λ_i and vectors $X_i=(x_i^{(1)}, \cdots, x_i^{(n)})$ satisfying (3) are called the *characteristic roots* and *characteristic column vectors* of the matrix A. (Note that A is necessarily square.) Also any row vector Y such that

$$YA = \lambda Y \tag{5}$$

for some value of λ is a *characteristic row vector* of A. Clearly the values of λ for which (5) is satisfied must be the same as before, for (2) must hold in both cases.

In the following we shall suppose *either that A has n distinct roots, or that to each root of multiplicity r there correspond r linearly independent characteristic vectors of each type* (i.e., A has linear elementary divisors), as is almost always the case in practice.* The sets of characteristic roots and corresponding vectors will be denoted by

* For a discussion of the case of non-linear elementary divisors, see the paper mentioned in Footnote 2.

$\lambda_1, \lambda_2, \cdots, \lambda_n$; $X_1, X_2, \cdots, X_n$; $Y_1, Y_2, \cdots, Y_n$. If any root λ_i is simple, the vectors X_i and Y_i are uniquely determined except for a constant factor; if two roots are equal, any linear combination of two corresponding characteristic vectors is a characteristic vector. In the following we shall understand by λ_i, X_i, X_i, $(i=1, 2, \cdots, n)$ the ith root and corresponding column and row vectors of some given matrix A of order n.

Since $(\lambda_j Y_j)X_i = Y_j A X_i = Y_j(\lambda_i X_i)$, we have $(\lambda_i - \lambda_j) Y_j X_i = 0$, and it follows that the characteristic column vector X_i is orthogonal to the characteristic row vector Y_j if $\lambda_i \neq \lambda_j$. We may clearly choose the vectors corresponding to any multiple roots in such a way that $Y_j X_i = 0$ whenever $i \neq j$, and we shall suppose this done in future.

In what follow, unless stated otherwise, we shall assume that *all the roots are distinct* and shall take λ_1 to be the root of *largest absolute value*.

2. Some important relations. Since the X_i's are linearly independent, any column vector V of order n may be expressed in the form

$$V = a_1 X_1 + a_2 X_2 + \cdots + a_n X_n, \tag{6}$$

where the a_i's are constants. Multiplying (2) by A, we get

$$AV = a_1 A X_1 + \cdots + a_n A X_n = a_1 \lambda_1 X_1 + \cdots + a_n \lambda_n X_n, \tag{7}$$

and by induction

$$A^k V = a_1 \lambda_1^k X_1 + a_2 \lambda_2^k X_2 + \cdots + a_n \lambda_n^k X_n \tag{8}$$

for any positive integer k. More generally, if $P(\lambda)$ is any polynomial in the variable λ, we have

$$P(A)V = a_1 P(\lambda_1) X_1 + a_2 P(\lambda_2) X_2 + \cdots + a_n P(\lambda_n) X_n. \tag{9}$$

By setting $V = X_i$, we see that the matrix $P(A)$ has the characteristic vectors $X_1, X_2, \cdots, X_n$, and the characteristic roots $P(\lambda_1), P(\lambda_2), \cdots, P(\lambda_n)$. (The same relations, of course, hold for row vectors.)

A further useful relation may be derived from equation (2), viz.,

$$a_{11} + a_{22} + \cdots + a_{nn} = \lambda_1 + \lambda_2 + \cdots + \lambda_n, \tag{10}$$

i.e., the sum of the roots is equal to the sum of the elements of the principal diagonal.

3. Symmetric matrices. A square matrix is *symmetric* if it is unchanged by turning its rows into columns and vice versa. The characteristic row vectors of a symmetric matrix have the same elements as the corresponding column vectors ($Y_i = X_i'$ for $i=1, 2, \cdots, n$); hence the characteristic vectors may be regarded as a *single mutually orthogonal set.*

It follows that all the roots of a symmetric matrix are *real*, for two conjugate complex roots would correspond to two characteristic vectors having conjugate complex elements, and such vectors could not be orthogonal.

II. The Method of Iteration

We now take up the problem of finding the characteristic roots and vectors of a given matrix A. For the sake of simplicity, we shall confine ourselves at first to the case where A is symmetric.

4. Expansion of the determinant. The most obvious procedure for solution would be to expand the determinant in Eq. (2) into a polynomial and then solve the resulting algebraic equation for λ. This method is sometimes the most expeditious, especially if use is made of the techniques of expansion described by Wayland.[4] Nevertheless it is necessary to consider other methods, for the following reasons.

(a) Solving (2) gives only the roots; if the characteristic vectors are desired, they must be found by a separate process.

(b) Finding the complex roots of an algebraic equation may be as difficult as finding those of a matrix directly.

(c) In case approximate values of the roots are known, the work of expanding the determinant will not be affected, though that of solving the algebraic equation may be. On the other hand, we shall see in Sec. IV that such information can be used to great advantage when the problem is solved by iteration.

(d) The methods to be discussed are better adapted to machine calculations.

We therefore take up direct methods based on the process of iteration.

5. Iteration. This method consists in multiplying a suitable vector V repeatedly by A. Representing V in the form (6), and considering equation (8), we see that if $a_1 \neq 0$ (i.e., if V is not orthogonal to X) the first term on the right will predominate more and more as k increases. In fact

$$A^k V = \lambda_1^k (a_1 X_1 + O(\lambda_2/\lambda_1)^k), \tag{11}$$

where λ_2 is the root of second highest absolute value. Similarly

$$A^{k+1} V = \lambda_1^{k+1} (a_1 X_1 + O(\lambda_2/\lambda_1)^{k+1}).$$

Thus *the ratio of corresponding elements of $A^k V$ and $A^{k+1} V$ approaches λ_1 as $k \to \infty$.* And if $A^k V$ is divided by one of its elements, *the resulting vector tends to a multiple of* X_1 (which, of course, is determined only up to a constant factor anyway). These points will be clarified by the following concrete case.

Example 3.

$$A = \left[\begin{array}{cccc|c} 6 & 1 & -1 & 3 & 9 \\ 1 & 4 & 0 & -2 & 3 \\ -1 & 0 & 1 & 5 & 5 \\ 3 & -2 & 5 & 2 & 8 \end{array}\right].$$

(The extra column on the right is the sum of the others; its purpose will be explained in the following.)

Choosing

$$V = (4, -1, 2, 4),$$

(more will be said later about the method of choice), we obtain

$$AV = (33, -8, 18, 32),$$
$$A^2 V = (268, -63, 145, 269),$$
$$A^3 V = (2207, -522, 1222, 2193),$$
$$A^4 V = (18077, -4267, 9980, 18161).$$

Successive ratios of the leading elements are

$$8.25,\ 8.12,\ 8.235,\ 8.191.$$

(The true value of the root is 8.22557331.) The vectors obtained by dividing the foregoing by their leading elements are (with the true vector for comparison)

$$(1.000,\ -.242,\ .545,\ .970),$$
$$(1.000,\ -.235,\ .541,\ 1.004),$$
$$(1.000,\ -.237,\ .554,\ .994),$$
$$(1.000,\ -.235,\ .552,\ 1.005);$$
$$X_1 = (1.00000000,\ -.24073464,\ .55955487,\ 1.00862094).$$

6. Use of powers of the matrix. It will be noticed that the convergence is not very rapid, and the multiplication might have to be repeated many times if λ_1 or the components of X_1 were wanted to many figures. The process may be shortened by first squaring A repeatedly:

$$A^2 = \left[\begin{array}{cccc|c} 47 & 4 & 8 & 17 & 76 \\ 14 & 21 & -11 & -9 & 5 \\ 8 & -11 & 27 & 12 & 36 \\ 17 & -9 & 12 & 42 & 62 \end{array}\right],$$

$$A^4 = \left[\begin{array}{cccc|c} 2578 & 31 & 752 & 1573 & 4934 \\ 31 & 659 & -604 & -631 & -545 \\ 752 & -604 & 1058 & 1063 & 2269 \\ 1573 & -631 & 1063 & 2278 & 4283 \end{array}\right],$$

$$A^8 = \cdots.$$

(The powers of a symmetric matrix are of course symmetric also.)

The *sum columns* are useful as checks at this point, for the sum column of A^2 should be the same as the result of operating with A on its own sum column, and similarly for the higher powers. That this is so can be seen by noting that the sum column of A is equal to AH, where $H=(1, 1, 1, 1)$. Then $A^2H=A(AH)$, $A^4H=A^2(A^2H)$, etc.

We can now multiply V directly by some higher power of A and thus save many stepts in the iteration. For example

$$A^4(A^4V) = A^8V = (82542442,\ -19739077,\ 46035155,\ 83107096)$$
$$A(A^8V) = A^9V = (678801708,\ -162628058,\ 379028193,\ 683495447)$$
$$A(A^9V) = A^2(A^8V) = A^{10}V = (5581640338,\ -1338701418,\ 311703720,\ 5623793099)$$

These are proportional to

$$(1.000,\ -.2391,\ -.5577,\ 1.0068),\quad (1.000,\ -.2396,\ .5584,\ 1.0069),$$
$$(1.0000,\ -.2398,\ .5586,\ 1.0076),$$

and the ratios of the leading elements are

$$8.224, \quad 8.224.$$

Note, however, that forming each power of A requires about as much computation as n vectors A^kV, so this procedure should not be carried too far.

For example, to compute $A^{16}V$ by operating directly with A on V sixteen times requires $16n$ multiplications; forming A^2 first and then operating eight times on V requires $\frac{1}{2}n(n+1)+8n$ multiplications. Similarly, using A^4 requires $n(n+1)+4n$; using A^8, $\frac{3}{2}n(n+1)+2n$; and using A^{16}, $2n(n+1)+n$. For $n=4$, it will be seen that working with A^4 is the most economical.

7. Finding the smaller roots. So far we have obtained only λ_1. What about the other roots?

Clearly these can be found if we can determine the products $\lambda_1\lambda_2$, $\lambda_1\lambda_2\lambda_3$, etc. We shall now see how to modify the iteration process in order to do this.

Suppose we have two sequences having kth terms of the form

$$\begin{aligned} \beta_k &\equiv b_1\lambda_1^k + b_2\lambda_2^k + \cdots + b_n\lambda_n^k \\ \beta_k' &\equiv b_1'\lambda_1^k + b_2'\lambda_2^k + \cdots + b_n'\lambda_n^k \end{aligned} \qquad (k = 0, 1, 2, \cdots). \tag{12}$$

Then we define

$$\begin{aligned} \beta_{2,k} &= \begin{vmatrix} \beta_k & \beta_{k+1} \\ \beta_k' & \beta_{k+1} \end{vmatrix} \\ &= \lambda_1^k\lambda_2^k(b_2b_1' - b_1b_2')(\lambda_1 - \lambda_2) + \lambda_1^k\lambda_3^k(b_3b_1' - b_1b_3')(\lambda_1 - \lambda_3) + \cdots, \end{aligned} \tag{13}$$

whence

$$\beta_{2,k+1}/\beta_{2,k} = \lambda_1\lambda_2 + O(\lambda_3/\lambda_2)^k \qquad (k = 0, 1, 2 \cdots). \tag{14}$$

To be concrete, if we have a sequence of vectors of the type V, AV, A^2V, $\cdots$ A^kV, $\cdots$ we could write $\beta_k=(A^kV)^{(1)}$, $\beta_k'=(A^kV)^{(2)}$, where $(A^kV)^{(i)}$ is the ith element of $A^kV(i=1, \cdots, n)$, so that

$$\beta_{2,k} = \begin{vmatrix} (A^kV)^{(1)} & (A^{k+1}V)^{(1)} \\ (A^kV)^{(2)} & (A^{k+1}V)^{(2)} \end{vmatrix}$$

(of course, any other pair of elements would work equally well). Consider the following case.

Example 2.

$$A = \begin{bmatrix} 0 & 0 & 2 & 1 \\ 0 & 1 & -3 & 1 \\ 2 & -3 & -3 & 4 \\ -1 & 1 & 4 & -2 \end{bmatrix}.$$

It is found that $\lambda_1 = -8.075320860$. Operating with A upon the vector V below yields

$$V = (5, -9, -2, 8),$$
$$AV = (-12, 5, 75, -38),$$
$$A^2V = (188, -258, -417, 393),$$
$$A^3V = (-225, 1383, 3970, -2896),$$
$$A^4V = (10836, -13423, -30093, 24280),$$
$$A^5V = (-84466, 101136, 249340, -193191).$$

Then

$$\begin{vmatrix} -2 & 75 \\ 8 & -38 \end{vmatrix} = -524, \qquad \begin{vmatrix} 75 & -416 \\ -38 & 393 \end{vmatrix} = 13667, \text{ etc.}$$

Values of successive determinants and their ratios are

$$-524 \qquad 13667 \qquad -355474 \qquad 9242272 \qquad -240278437$$
$$-26.08 \qquad -26.01 \qquad -25.9985 \qquad -25.9978$$

Thus $\lambda_1\lambda_2 = -26.00$ and $\lambda_2 = 3.219$.

On the other hand, we may take $\beta_k' = \beta_{k+1}$, in which case

$$\beta_{2,k} = \begin{vmatrix} \beta_k & \beta_{k+1} \\ \beta_{k+1} & \beta_{k+2} \end{vmatrix}. \tag{15}$$

That is, we can set up a sequence converging to $\lambda_1\lambda_2$ given any single sequence of the form (12)

If we wish to find λ_3 ,we select three sequences of the form (12), set up the corresponding three-rowed determinant $\beta_{3,k}$, and get as before

$$\beta_{3,k+1}/\beta_{3,k} = \lambda_1\lambda_2\lambda_3 + O(\lambda_4/\lambda_3)^k. \tag{16}$$

Note however, that the characteristic vectors $X_2, X_3, \cdots, X_n$ cannot be found by this means.

III. Devices for Improving Convergence

If we employ the method of iteration in the crude form, described above, we find that the convergence often leaves much to be desired in terms of time consumed. In this section we consider various ways of improving the efficiency of the method.

8. Scalar products. Consider the scalar product (see Sec. I, 1) $(V'A^k)\cdot(A^lV)$, $(k, l = 0, 1, 2, 3, \cdots)$.

This may be written (in view of (6) and the orthogonality of X_i's)

$$V'A^kV^lV = V'A^{k+l}V = a_1\lambda_1^{k+l}X_1'X_1 + a_2\lambda_2^{k+l}X_2'X_2 + \cdots + a_n\lambda_n^{k+l}X_n'X_n. \tag{17}$$

In particular

$$V'A^{k-1}\cdot A^kV = a_1\lambda_1^{2k-1}X_1'X_1 + a_2\lambda_2^{2k-1}X_2'X_2 + \cdots + a_n\lambda_n^{2k-1}X_n'X_n,$$

and

$$V'A^k\cdot A^kV = a_1\lambda_1^{2k}X_1'X_1 + a_2\lambda_2^{2k}X_2'X_2 + \cdots + a_n\lambda_n^{2k}X_n'X_n,$$

so that

$$\frac{V'A^k \cdot A^k V}{V'A^{k-1} \cdot A^k V} = \lambda_1 + O(\lambda_2/\lambda_1)^{2k-1}, \qquad (k = 1, 2, 3, \cdots). \tag{18}$$

It follows from (18) that by taking suitable scalar products the number of steps required to find λ_1 to a given degree of accuracy is *approximately halved.*

To illustrate this method, let us apply it to the vector sequence of Example 1. Working with the last two vectors, we get

$$V'A^{18}V = 1098048041, \qquad V'A^{19}V/V'A^{18}V = 8.2255708,$$
$$V'A^{19}V = 9032871992, \qquad V'A^{20}V/V'A^{19}V = 8.2255717.$$
$$V'A^{20}V = 74293955655,$$

(The last figures of the scalar products have been omitted.) The improvement in the approximation is evident.

The method can also be applied to finding the products of the roots, by making proper use of (15).

9. Choice of initial vector. So far nothing has been said about the choice of the initial vector V. Clearly, the best V is one that is already "near" to X_1, i.e., one having a_1 (equation (6)) large in comparison with the other a's. A few suggestions may be helpful. The sum columns of A and even more of A^2, A^3, etc., may be good indications, though sometimes they are misleading, as when H happens to be nearly orthogonal to X_1; in this case we may be led astray to one of the other vectors. The other columns of A, especially those having the largest elements, are also good first guesses. In any case, the first V's should have small integers for elements and should be tested by trial multiplications by A.

10. Reduction of leading element to unity. Once V has been selected and the iteration is under way, it is clear (since only the *ratios* of the vector elements are significant) that the approximation to X_1 is not affected if each vector A^kV is divided through by one of its elements—say the first if this is not zero. Then it will be clear from the extent of agreement of successive vectors how far the approximation has progressed, and time can be saved by retaining in the elements only as many digits as appear to be correct. (The choice of V is actually slightly altered at each step, but this is of no importance.) The following example shows the procedure.

Example 3.

$$A = \begin{bmatrix} -2 & -2 & 0 & 3 & -1 \\ -2 & 0 & -3 & 5 & 0 \\ 0 & -3 & -5 & 11 & 1 \\ 3 & 5 & 1 & -3 & -1 \\ -1 & 0 & 1 & -1 & -1 \end{bmatrix}.$$

Casual inspection suggests the vector

$$V = (1, 1, 1, -1, 0)$$

as a first approximation of X_1, The vector AV resembles V sufficiently to indicate

that we are on the right track. Operating repeatedly with A and dividing through by the leading element each time yields the sequence below:

$$\begin{aligned} AV &= (-7 \ \ -10, \ -9, \ 12, \ 1)(69), \\ &\quad (1.00, \ 1.46, \ 1.28, \ -1.70, \ -.22)(-9.80), \\ &\quad (1.000, \ 1.463, \ 1.296, \ -1.724, \ -.224)(-9.874), \\ &\quad (1.000, \ 1.469, \ 1.298, \ -1.722, \ -.227)(-9.877), \\ &\quad (1.000, \ 1.468, \ 1.301, \ -1.725, \ -.227)(-9.884); \\ X_1 &= (1.000000, \ 1.469801, \ 1.302061, \ -1.724997, \ -.228106). \end{aligned}$$

The figure in parentheses after each vector is the number by which the elements of the next vector have to be divided to reduce its leading element to unity. These numbers converge, of course, to λ_1. Note that while computing errors may slow up the convergence, they will not invalidate the final result.

To obtain a more accurate value of λ_1, we should operate with A on the last vector in the sequence and take scalar products as before. The number of correct significant figures thus obtained should be about double that of the last number in parentheses.

11. Aitken's δ^2 process. Since the rapidity of convergence in all the above processes depends essentially on the ratio $|\lambda_2/\lambda_1|$, where λ_2 is the root of second largest absolute value, the convergence is slow if this ratio is near 1. In such cases the convergence may be speeded up by using the δ^2 process, due to Aitken, which we now explain.

Our approximating sequences are all of the type

$$\phi_k = \lambda_1 + c_1 q_1^k + c_2 q_2^k + \cdots \qquad (k = 0, 1, 2, \cdots ; 1 > |q_1| \geqq |q_2| \geqq \cdots) \tag{19}$$

where the q's are of the form λ_2/λ_1, λ_3/λ_1, $(\lambda_2/\lambda_1)^2$, etc.

Writing

$$\psi_{2,k} = \begin{vmatrix} \phi_k & \phi_{k+1} \\ \phi_{k+1} & \phi_{k+2} \end{vmatrix}, \qquad (k = 0, 1, 2, \cdots), \tag{20}$$

we readily see that

$$\psi_{2,k} = \lambda_1[c_1(1 - q_1)^2 q_1^k + c_2(1 - q_2)^2 q_2^k + \cdots] + c_1 c_2 q_1^k q_2^k (q_1 - q_2)^2 + \cdots .$$

On the other hand

$$\delta^2\phi_{k+1} \equiv \phi_{k+2} - 2\phi_{k+1} + \phi_k = c_1(1 - q_1)^2 q_1^k + c_2(1 - q_2)^2 q_2^k + \cdots .$$

Defining

$$\phi_{2,k} = \psi_{2,k}/\delta^2\phi_{k+1} \qquad (k = 0, 1, 2, \cdots), \tag{21}$$

we see that

$$\phi_{2,k} = \lambda_1 + O(q_2^k). \tag{22}$$

Thus, if $|q_2| < |q_1|$, the $\phi_{2,k}$*'s will form a sequence converging to* λ_1 *more rapidly than the sequence of* ϕ_k*'s.* Applying the same process to the $\phi_{2,k}$'s will then result in a still more rapidly convergent sequence, and so on.

However, the remainder term in the sequence obtained by applying the δ^2 process twice is of the form $O[(q_2^2/q_1)^k]+O(q_3^k)$. Thus the improvement in convergence will be slight if $|q_1|$ and $|q_2|$ are nearly equal, especially if $q_2 \sim -q_1$.

12. Extension of the δ^2 process. For this case, we shall develop a modified procedure which proves more efficacious.

We define a *second derived sequence*

$$\phi_{3,k} = \psi_{3,k}/\chi_{3,k} \qquad (k = 0, 1, 2, \cdots), \tag{23}$$

where

$$\psi_{3,k} = \begin{vmatrix} \phi_k & \phi_{k+1} & \phi_{k+2} \\ \phi_{k+1} & \phi_{k+2} & \phi_{k+3} \\ \phi_{k+2} & \phi_{k+3} & \phi_{k+4} \end{vmatrix} = \begin{vmatrix} \psi_{2,k+1} & \psi_{2,k+2} \\ \psi_{2,k+2} & \psi_{2,k+3} \end{vmatrix} \Big/ \phi_{k+2} \tag{24}$$

and

$$\chi_{3,k} = \begin{vmatrix} \delta^2\phi_{k+1} & \delta^2\phi_{k+2} \\ \delta^2\phi_{k+2} & \delta^2\phi_{k+3} \end{vmatrix}. \tag{25}$$

It can be verified that

$$\phi_{3,k} = \lambda_1 + O(q_3^k). \tag{26}$$

It should be noted that if the elements ϕ_k of the original sequence are replaced by $\phi_k - c$, where c is a constant, all the sequences will converge to $\lambda_1 - c$. Thus if the first few digits of the ϕ_k's are the same for all k, we may simply ignore these digits in computing the derived sequences.

Example 4.

$$A = \begin{bmatrix} 2 & 1 & -1 & 3 \\ 1 & 0 & 0 & -2 \\ -1 & 0 & -3 & 5 \\ 3 & -2 & 5 & -2 \end{bmatrix} \qquad V = \begin{vmatrix} 1 \\ -1 \\ 3 \\ -3 \end{vmatrix}.$$

Carrying out the foregoing procedure, we form the following table:

$V'A^kV$	ϕ_k	$\delta^2\phi_k$	$\psi_{2,k-1}$	$\phi_{2,k-1}$	$\phi_{3,k-2}$
20	−8.550000000			−8.59	
−171	−8.602339181	.066643012	−.006071541	1105433	−8.59
1471	−8.588035350	−.018447694	.001683322	1248386	120357
−12633	−8.592179213	.005452930	−.000497222	1184409	120331
108545	−8.590870146	−.001769071	.000161358	1210537	
−932496	−8.591330150				
8011381					

(In the above, $\phi_k - 8.5$, rather than ϕ_k, is used in computing the later columns.)

Further derived sequences, showing still better convergence, could be defined in a similar fashion. However, the extra computation required and the rapid accumulation of rounding-off errors make such a procedure of doubtful value.

The same process can be used in finding the elements of X_1 (see the paper quoted in Footnote 2).

IV. Finding the Remaining Roots and Vectors

In Secs. II and III we saw how all the *roots* of A could be obtained, though with limited accuracy for the smaller roots in practice. Now we consider methods of getting both roots and vectors at once, at the same time improving the accuracy of the former.

13. λ_1-differencing. Perhaps the simplest method is that of λ_1-differencing the sequence (8). That is, we form the differences

$$A^{k+1}V - \lambda_1 A^k V = a_2(\lambda_2 - \lambda_1)\lambda_2^k X_2 + a_3(\lambda_3 - \lambda_1)\lambda_3^k X_3 + \cdots + a_n(\lambda_n - \lambda_1)\lambda_n^k X_n. \quad (27)$$

If λ_2 is larger numerically than the other roots and $a_2 \neq 0$ the first term predominates, and we find X_2 and λ_2 much as we originally found X_1 and λ_1.

The disadvantage of this method is that nearly equal numbers have to be subtracted at each stage with a resulting loss of significant figures. The better the convergence of the original sequence is, the worse this becomes. Needless to say, the difficulty is increased if we try to find X_3, X_4, etc., by repeating the procedure.

14. Deflation.[1,2] We may also proceed by *deflation*, which consists in replacing the matrix A by a matrix A_1 having the same characteristic roots and vectors as A, except that λ_1 is replaced by zero. Applying the method of iteration to A_1 will then yield λ_2 and X_2. Deflating again give λ_3 and X_3, and so on.

We set up the matrix A_1 as follows:

$$A_1 = A - \frac{\lambda_1}{X_1' X_1} X_1 X_1'. \quad (28)$$

Note that $X_1' X_1$ is the *scalar* product (sum of squares) while $X_1 X_1'$ is the *matrix* product

$$\begin{bmatrix} x_1^{(1)2} & x_1^{(1)}x_1^{(2)} & \cdots & x_1^{(1)}x_1^{(n)} \\ x_1^{(2)}x_1^{(1)} & x_1^{(2)2} & \cdots & x_1^{(2)}x_1^{(n)} \\ \cdots & \cdots & \cdots & \cdots \\ x_1^{(n)}x_1^{(1)} & x_1^{(n)}x_1^{(2)} & \cdots & x_1^{(n)2} \end{bmatrix}.$$

That A_1 has the desired properties can be seen by substitution:

$$A_1 X_1 = AX_1 - \frac{\lambda_1}{X_1' X_1} X_1 X_1' X_1 = \lambda_1 X_1 - \lambda_1 X_1 = 0,$$

$$A_1 X_i = AX_i - \frac{\lambda_1}{X_1' X_1} X_1 X_1' X_i = \lambda_1 X_i \qquad (i \neq 1).$$

Let us apply this procedure to Example 1:

$$X_1' X_1 = 2.38837102, \qquad \lambda_1 / X_1' X_1 = 3.44400985$$

$$A_1 = \left[\begin{array}{cccc|c} 2.55599015 & 1.82909247 & -2.92711248 & -.47370045 & .98426969 \\ 1.82909247 & 3.71445974 & .46392273 & -1.16375997 & 4.84371497 \\ -2.92711248 & .46392273 & -.07832518 & 3.05627400 & .51475907 \\ -.47370045 & -1.16375997 & 3.05627400 & -1.50364702 & -.08483344 \end{array}\right].$$

Starting with the vector V_1 below and proceeding as in Sec. III, 10, we get in succession the vectors

$$\begin{aligned} V_1 &= (2, 2, -1, -1)(12.216), \\ &\quad (1.00, 0.96, -0.65, -0.40)(6.4040), \\ &\quad (1.00, 0.87, -0.57, -0.46)(6.0337), \\ &\quad (1.00, 0.88, -0.64, -0.42)(6.2379), \\ &\quad (1.00, 0.85, -0.60, -0.45)(6.0802) \end{aligned}$$

whereas the vector and root to four places are

$$X_2 = (1.0000, 0.8681, -0.6198, -0.4404), \lambda_2 = 0.16666.$$

We see that the difficulty of losing significant figures by subtraction is largely avoided, while the convergence toward X_2 is fairly rapid. On the other hand, the extra figures in the elements of A_1 make the further computations somewhat laborious, and in any case the accuracy attainable in X_2 is limited by that of λ_1 and X_1.

Some saving of time may be affected by noting that

$$A_1^k = A^k - \frac{\lambda_1^k}{X_1' X_1} X_1 X_1',$$

whence

$$A_1^k V = A^k V = \frac{\lambda_1^k}{X_1' X_1} X_1 X_1' V,$$

i.e., the vector $A_1^k V$ can be obtained from $A^k V$ without setting up the matrix A_1. But this again brings a loss of significant figures; thus this device is of limited applicability.

15. Use of matrix polynomials. In view of what has been said above, it would seem that there is room for a different approach. Now, upon reviewing our work up to this point, it will be seen that we have tacitly assumed that we had no prior notion of the location of the roots of our matrix A, and we have indicated no way of using such information if we had it. However, in a practical case growing out of, say, a vibration problem, we could usually make at least a guess at where the roots lie. In any event, we could get an approximation to the largest root by a couple of steps of iteration, and we shall see shortly how the other roots might be estimated.

Suppose, then, that we have approximate values $\lambda_1', \lambda_2', \cdots, \lambda_n'$ of the n roots $\lambda_1, \lambda_2, \cdots, \lambda_n$. Now consider the polynomial.

$$P_1(\lambda) \equiv (\lambda - \lambda_2')(\lambda - \lambda_3') \cdots (\lambda - \lambda_n') \tag{28}$$

and the corresponding *matrix*

$$P(A) = (A - \lambda_2')(A - \lambda_3') \cdots (A - \lambda_n'). \tag{29}$$

It follows from equation (9) that $P(A)$ has the characteristic roots $P_1(\lambda_1)$, $P_1(\lambda_2), \cdots, P_1(\lambda_n)$. But by definition

$$P_1(\lambda_2') = P_1(\lambda_3') = \cdots = P_1(\lambda_n') = 0,$$

and since $P(\lambda)$ is continuous one would expect $P_1(\lambda_2)$, $P_1(\lambda_3)$, $\cdots$, $P_1(\lambda_n)$ to be small, how small depending on how good the approximations were. On the other hand, the roots being supposed distinct, $P_1(\lambda_1)$ would not in general be near zero, and therefore

$$|P_1(\lambda_1)| \gg |P_1(\lambda_i)|, \qquad i \neq 1.$$

If we now choose a vector V of the form (6), with $a_1 \neq 0$, we have by (9)

$$P_1(A)\cdot V = a_1P_1(\lambda_1)X_1 + a_2P_1(\lambda_2)X_2 + \cdots + a_nP_1(\lambda_n)X_n. \tag{30}$$

The first term on the right will predominate strongly, so that if we proceed by iteration, i.e., multiply repeatedly by $P_1(A)$, *the convergence to* X_1 *will be very rapid.* (Of course it is all the better if V is already an approximation to X_1.)

Once X_1 is obtained, λ_1 can be determined to the same degree of accuracy by multiplying X_1 by A, and to greater accuracy by forming scalar products.

Note that nothing is really altered if λ_1 is replaced by one of the other roots. For example, to find λ_2 and X_2 we could use

$$P_2(A) \equiv (A - \lambda_1')(A - \lambda_3')(A - \lambda_4') \cdots (A - \lambda_n').$$

Example 5. Suppose that for a given fourth-order matrix we have

$$\lambda_1 \sim 6.1 = \lambda_1', \quad \lambda_2 \sim -4.3 = \lambda_2', \quad \lambda_3 \sim -2.6 = \lambda_3', \quad \lambda_4 \sim 0.4 = \lambda_4'.$$

Then

$$P_1(\lambda) = (\lambda + 4.3)(\lambda - 2.6)(\lambda - 0.4) = \lambda^3 + 1.3\lambda^2 - 11.9\lambda + 4.5,$$

where the coefficients have been rounded off to one decimal place, as there is clearly no advantage in retaining more. In fact, we should in practice make a further simplification, and replace $P_1(\lambda)$ as written above by a polynomial like $\lambda^3+\lambda^2-12\lambda+4$ or $2\lambda^3+3\lambda^2-24\lambda+9$, whose coefficients are *small integers.* In this way, the subsequent computations would be simplified without slowing down the convergence much. Where to strike the balance in practice is largely a matter of experience and "feel" on the part of the computer.

16. Application of orthogonality. Let us return to Sec. I, 2, but suppose that V, instead of being a completely arbitrary vector, is *orthogonal to* X_1, i.e., $a_1 = 0$. (We shall see later how to select such a V.) Then

$$\begin{aligned} A^kV &= a_2\lambda_2^kX_2 + a_3\lambda_3^kX_3 + \cdots + a_n\lambda_n^kX_n, \\ P(A)V &= a_2P(\lambda_2)V_2 + a_3P(\lambda_3)X_3 + \cdots + a_nP(\lambda_n)X_n. \end{aligned} \tag{31}$$

To see the application of this, consider Example 5 again and suppose that X_1 has been found to the desired accuracy. From what has been said before, our first step in gett X_2 would be to set up the polynomial

$$P_2(\lambda) = (\lambda - 6.1)(\lambda - 2.6)(\lambda - 0.4).$$

However, if we are to work with a vector V orthogonal to X_1, we might just as well use the simpler polynomial

$$Q_2(\lambda) = (\lambda - 2.6)(\lambda - 0.4).$$

The fact that $Q_2(\lambda_1)$ is not small compared with $Q_2(\lambda_2)$ has no effect on the iteration, for

$$Q_2(A)\cdot V = a_2Q_2(\lambda_2)X_2 + a_3Q_2(\lambda_3)X_3 + a_4Q_2(\lambda_4)X_4.$$

How can we select such a vector V? Consider any vector $V_0=(v_0^{(1)}, v_0^{(2)}, \cdots, v_0^{(n)})$ Then *the vector*

$$\left(v_0^{(1)}, v_0^{(2)}, \cdots, v_0^{(n-1)} - \frac{v_0^{(1)}x_1^{(1)} + v_0^{(2)}x_1^{(2)} + \cdots + v_0^{(n-1)}x_1^{(n-1)}}{x_1^{(n)}}\right) \tag{32}$$

will be orthogonal to X_1 and can serve as V (provided $x_1^{(n)}\neq 0$).

Note that we have altered only the last element $v_0^{(n)}$ of V_0. We could, of course, have altered any other element $v_0^{(i)}$ instead, provided $x_1^{(i)}\neq 0$. In general it is best to alter an element corresponding to one of the larger elements of X_1 so that V may differ as little as possible from V_0.

In order to avoid the accumulation of rounding-off errors, it is generally desirable to repeat the orthogonalization at each step of the iteration. Also, it is well to determine the elements of X_1 to one more significant figure than is required in those of the other vectors. (Of course we could use any of the vectors X_i as a starting point.)

17. Obtaining the first approximations. It is well to give some attention to the question of getting rough values of the roots when these are not available beforehand. Since the procedure is rough and ready rather than fixed, it is best shown by an example. For this purpose we return to the matrix A of Example 3.

We have already seen that

$$\lambda_1 = -9.886,\ X_1 = (1.000,\ 1.470,\ 1.302,\ -1.725,\ -0.228).$$

Looking back, we see that one or two steps in the iteration would have sufficed to give us one-place accuracy, which is all we need at this stage.

Referring to equation (10), we see that

$$\lambda_1+\lambda_2+\lambda_3+\lambda_4+\lambda_5 = -2+0-5-3-1 = -11, \tag{33}$$

so that A must certainly have a root λ_2 greater than -1. It follows that the matrix $A_1=A+6$ will have a root $\lambda_2+6>5$, which is greater in absolute value than its root λ_1+6. Trying out simple vectors like (1, 0, 0, 0, 0), (0, 1, 0, 0, 0), etc., we see that the combination

$$V_1 = (0,\ 1,\ 0,\ 1,\ 0)$$

looks hopeful. In fact

$$A_1V_1 = (1,\ 11,\ -2,\ 8,\ -1), \qquad A_1^2V = (7,\ 110,\ -28,\ 81,\ -16),$$

and by taking scalar products we get $\lambda_2+6\sim 10.2$, $\lambda_2\sim 4.2$, while $A_1^2V_1$ will stand as a first approximation to X_2.

To proceed further, we note that $\lambda_1+3\sim -6.9$, $\lambda_2+3\sim 7.2$. Therefore the matrix $(A+3)^2-50$ will have two roots near 0 and three between -50 and 0. This matrix could be used to obtain another root λ_3 of A. However, it probably has two or three roots near -50, which would be difficult to separate, so we used instead the matrix $A_2=(A^2+6A+9)-35=A(A+6)-26$.

After a few trials we decide on the vector $V_2=(2, 0, 0, 1, 4)$. Then $A_2V_2=(-70, -4, -6, -27, -128)$, and taking scalar products yields

$$(\lambda_3+3)^2-35\sim-32.5, \qquad (\lambda_3+3)^2\sim 2.5, \qquad \lambda_3+3\sim\pm 1.6,$$

so that $\lambda_3\sim-1.4$ or -4.6. Since $AV_2=(-5, 1, 5, -1, -7)$ and $A_1V_2=(7, 1, 5, 5, 17)$, the value $\lambda_3\sim-1.4$ must be the correct one.

To get approximations to λ_4 and λ_5, we take the sum of the principal diagonal terms in A_2:

$$\sum_{i=1}^{5}(\lambda_i^2+6\lambda_i-26)=-55,$$

so that, from (33)

$$\lambda_1^2+\lambda_2^2+\lambda_3^2+\lambda_4^2+\lambda_5^2=141. \tag{34}$$

Combining (33) and (34) with the known values of λ_1, λ_2, and λ_3, we get $\lambda_4\sim0.9$, $\lambda_5\sim-4.8$.

18. Examples. Two more examples show how to carry through the process.

Example 6. Find to six decimal places the roots of

$$A=\left[\begin{array}{cccc|c} 2 & 2 & 0 & 4 & 8 \\ 2 & -1 & -1 & 3 & 3 \\ 0 & -1 & 0 & -2 & -3 \\ 4 & 3 & 2 & 0 & 5 \end{array}\right].$$

Using the methods of the last example, we conclude that $\lambda_1\sim7.06$, $\lambda_2\sim-4.19$, $\lambda_3\sim0.45$, $\lambda_4\sim-2.32$, $X_1\sim(106, 67, -38, 100)$, $X_2\sim(53, 53, -41, -106)=V_2$. We next compute X_1 to greater accuracy in order to obtain λ_1 to six places. Since we want λ_1 to seven significant figures, four figures in the elements of X_1 should be sufficient, as the formation of scalar products about doubles the number of significant figures, as explained above. In order to avoid rounding-off errors, both here and in the later calculations, we carry an extra figure.

If we were to follow rigidly the procedure outline above (paragraph 18), we should begin by writing the polynomial $(\lambda+4.19)(\lambda-0.45)(\lambda+2.32)$. However, we note that the simple *linear* matrix polynomial $A_2=A_1+2$ has very favorable root ratios, and since the number of figures required is not great anyway, it seems worthwhile to go ahead with this matrix rather than taking the time to set up a more complicated expression. Applying A_2 to our first approximation and dividing by the leading element at each step yields successively

(106, 67, −38, 100)(958),

(1.000, .644, −.358, .941)(9.052),

(1.0000, .6435, −.3582, 19423)(9.0562),

(1.00000, .64360, −.35826, .96206)(9.05544),

(1.00000, .64360, −.35826, .94213)(9.05572),

(9.05572, 5.82825, −3.24438, 8.53158).

The last vector was obtained by applying A_2 once more without division. Taking scalar products of the last two vectors, we get

$$\begin{array}{ll} 2.4301801245 & \\ 22.0069707442 & 9.055695307 \\ 199.2884216617 & 9.055695307 \end{array}$$

We conclude that $\lambda_1 = 7.055695307$, to greater accuracy than was required.

We could now find λ_2 in the same manner by starting with $A-2$, say. However the convergence would be rather slow because the ratios of the roots would be less favorable than before. Instead, we proceed as in Example 5. Since

$$(\lambda + 2.32)(\lambda - .45) = \lambda^2 + 1.87\lambda - 1.04,$$

we set up the matrix $A_3 = A^2 + 2A - 1$. Now if we start with a vector orthogonal to X_1, we shall get a sequence converging to X_2. We therefore start with the vector V_2, but modify its last element so as to make it orthogonal to X_1. The modified last element is

$$-[53(1.00000) + 53(.64360) - 41(-.35826)]/.94213 = -108.05. \tag{35}$$

We now apply A_3 to the modified vector several times, repeating the orthogonalization process at each step. We get successively

$$(53, 53, -41, -108.05)(417.90),$$
$$(1.000, 1.130, -.739, -2.11438)(8.21364),$$
$$(1.0000, 1.1257, -.7384, -2.11122)(8.19976),$$
$$(1.00000, 1.12581, -.73843, -2.11130)(8.20028),$$
$$(1.00000, 1.12579, -.73842, -2.11128).$$

Applying $A_4 = A - 2$ (of which $\lambda_2 - 2$ is the numerically largest root) and taking scalar products yields $\lambda_2 - 2 = -6.1937207$, $\lambda_2 = -4.1937207$. It follows by direct calculation that $\lambda_3 = 0.464791$, $\lambda_4 = -2.326766$.

Example 6. Find to five significant figures the characteristic vectors of the matrix

$$B = \left[\begin{array}{cccc|c} 2 & 1 & 0 & 4 & 7 \\ 1 & -1 & -1 & 3 & 2 \\ 0 & -1 & 8 & -2 & -3 \\ 4 & 3 & -2 & 0 & 5 \end{array}\right]$$

First we compute B^2 and B^3.

Proceedings as in the previous example, we shortly conclude that $X_1 \sim (76, 44, -30, 76)$, $X_2 \sim (67, 84, -37, -129)$, $\lambda_1 \sim 6.56$, $\lambda_2 \sim -4.53$, $\lambda_3 \sim 0.69$, $\lambda_4 \sim -1.72$.

Next we find X_1 to the desired accuracy. Since more accuracy is required than in the previous example, it is worth while to go to more trouble to secure rapid convergence. We have

$$(\lambda - 0.69)(\lambda + 1.72)(\lambda + 4.53) = \lambda^3 + 5.56\lambda^2 + 3.48\lambda - 5.39.$$

Thus $B_1=2B^3+11B+7B-11$ will have three roots near zero. We get successively

(76, 44, −30, 76)(81506),

(1.0000000, .5789512, −.3916276, .9944053)(1069.511549),

(1.0000000, .5789483, −.3916329, .9944266)(1069.520900),

(1.0000000, .5789483, −.3916329, .9944265).

The rapidity of convergence is to be noted. (As before, the extra digits are retained to avoid rounding-off errors both here and in the orthogonalization that follows.)

To obtain the vector X_2, we write $(\lambda-0.69)(\lambda+1.72)=\lambda^2+1.03\lambda-1.19$, and set up the matrix $B_2=B^2+B-1$. As in the previous example, we must operate on vectors orthogonal to X_1. Orthogonalizing at each step we get in succession

(67, 84, −37, −129)(1048),

(1.0000, 1.1698, −.5821, −1.915902)(14.877570),

(1.0000, 1.1894, −.5893, −1.930148)(15.003080),

(1.000000, 1.189683, −.589403, −1.9303534)(15.0048880),

(1.000000, 1.189687, −.589404, −1.9303561).

The convergence could have been improved by setting up a cubic matrix polynomial as was done with X_1, but the extra time needed for doing so would have more than offset that saved later.

We could now obtain X_3 by starting with $B+2$, say, and working with vectors orthogonal to both X_1 and X_2. The double orthogonalization, however, would be time-consuming, so we start again with $(\lambda+1.72)(\lambda+4.53)=\lambda^2+6.25\lambda+7.79$.

Use of B^2+6B+8 suggests itself. However, a quick estimate of the roots indicates that this will not lead to very rapid convergence, and as yet we have not even a first approximation to X_3. So we use instead $B_3=4B^2+25B+31$.

Starting with a vector that turns out to be a rather poor guess and orthogonalizing with respect to X_1 at each step yields the following rapidly convergent sequence:

(1, 0, 0, −1),

(21, −10, 26, −5)(1036),

(1.0000, −.5097, 1.2838, −.203265)(50.266140),

(1.000000, −.516826, 1.293370, −.195347)(50.513110),

(1.000000, −.516805, 1.293339, −.1953714).

The vector X_4 can now be determined from the fact that it is orthogonal to X_1, X_2, and X_3. We set

$$X_4 = V + a_1X_1 + a_2X_2 + a_3X_3, \tag{36}$$

where V is any vector not orthogonal to X_1'. Multiplying (36) by X_1' gives $0=X_1'V+a_1X_1'X_1$, whence

$$a_1 = -X_1'V/X_1'X_1 \tag{37}$$

and similarly for a_2 and a_3. Choosing $V=(1, 0, 0, 0)$, we obtain $a_1=-0.4036424$, $a_2=-0.1541062$, $a_3=-0.3357978$. Substituting in (36) and dividing through by the leading element of the resulting vector yields $X_4=(1.000000, -2.287233, -1.741508, -0.359851)$.

The characteristic roots can be found, if desired, simply by computing the leading elements of BX_1, BX_2, BX_3, and BX_4. Greater accuracy can be achieved by forming scalar products as in previous examples.

19. Summary. We now summarize our procedure.

(a) Obtain rough values of all the roots. This may be done by operating on properly chosen vectors with matrices of the forms $A+n$, $(A+m)^2+n$, etc. After all but two of the roots have been located in this way, the last two can be found by solving a quadratic equation.

The examples indicate that three significant figures in the roots are amply sufficient at this stage, and one could often get along with fewer. Certainly a disproportionate amount of time should not be spent in getting these first approximations, as their sole purpose is the saving of time in the later steps. Fortunately the question is beside the point in many practical cases, where rough values of the roots are known to begin with; this step can then be omitted.

Naturally, if only a few figures are wanted in the final results, it may prove most economical to carry the work to a conclusion at this stage, rather than taking the time to set up the matrix polynomials of the succeeding steps.

(b) Select a polynomial $P(\lambda)$ such that, for some λ_i, $|P(\lambda_i)|>|P(\lambda_j)|$, $j\neq i$, and operate on a suitably chosen vector with $P(A)$ until X_i is obtained to the desired degree of accuracy.

$P(A)$ will ordinarily be chosen as in the examples. It will be noted that one or two figures in the coefficients are enough for rapid convergence, and more should not be used as time will be wasted in computing the elements of $P(A)$. It should be borne in mind that the coefficients of the higher powers are the most important.

Care should be taken to avoid errors at this stage, as much time can be lost if $P(A)$ is incorrectly chosen or written.

An extra place should be retained in the elements of X_i to avoid rounding-off errors, and it may be well to keep two such places if vectors orthogonal to X_i are to be constructed in step (*c*).

(c) Proceed similarly to obtain the other characteristic vectors, making use of the orthogonality of the vectors as in the examples given. If only the roots are desired, two of the vectors need not be computed. If all the vectors are wanted, the last one may be obtained from the others by the condition of orthogonality.

(d) Once the vectors are obtained, the corresponding roots may be found to about twice as many significant figures as the vectors by the scalar product method. In case the order of the matrix is greater than four, it may be necessary to use a matrix of the form $(A+m)^2+n$ to do this, getting first the square of the root and then finding the root itself by taking a square root. The last two roots may be obtained from the others by solving a quadratic as before.

In case A has multiple roots, this fact should become evident while carrying out step (a). We can then proceed much as before, except that we should not, of course, expect to find unique vectors corresponding to the multiple roots.

V. General Matrices; Complex Roots

We now approach the problem of the general (unsymmetric) matrix. Unlike the symmetric case, the characteristic roots are no longer necessarily real, and there are two sets of characteristic vectors instead of one, having less convenient orthogonality properties.

Most of the methods of the preceding section remain applicable. However, we cano no longer make such effective use of the orthogonality relation. Also, since the elements of the row vector $V'A$ need not be the same as those of the column vector AV, the method of scalar products no longer effects a great saving of time. On the other hand, new methods are required to handle complex roots.

20. Real roots. The case where all the roots are real presents few difficulties, and the procedure is very similar to that for the symmetric case. Some use may even be made of the orthogonality properties.

The same can be said if only two of the roots are complex, and the real roots can be found first, since the complex pair can then be obtained by solving a quadratic equation. The following example illustrates these points.

Example 7.

$$A = \left[\begin{array}{cccc|c} 2 & 0 & -1 & -3 & -2 \\ 1 & -3 & 0 & -2 & -4 \\ -2 & 1 & 2 & 1 & 2 \\ 3 & 4 & 0 & -1 & 6 \end{array}\right].$$

Find all the roots correct to six places, given that $\lambda_1 \sim 2.49$. $\lambda_2 \sim -1.77$.

We compute A^2 and A^3 and deduce that $\lambda_{3,4} \sim -0.36 \pm 3.28i$. Since

$$(\lambda + .36 - 3.28i)(\lambda + .36 + 3.28i)(\lambda + 1.77) = \lambda^3 + 2.45\lambda^2 + 12.2\lambda + 19.3,$$

we can get λ_1 by using the matrix $A_1 = 4A^3 + 10A + 49A + 77$. After a few steps we obtain

$$X_1 = (-.3637407, .0336918, 1.0000000, -.2743017),$$

whence (multiplying by A once) $\lambda_1 = 2.4868715$.

To find λ_2, we make use of the fact that the characteristic row vector Y_2 is orthogonal to X_1. We have

$$(\lambda + .36 - 3.28i)(\lambda + .36 + 3.28i) = \lambda^2 + .72\lambda + 10.9.$$

Writing

$$A_2 = 10A^2 + 7A + 109 = \left[\begin{array}{cccc|c} 93 & -130 & -47 & -61 & -145 \\ -63 & 98 & -10 & 36 & 61 \\ -54 & 37 & 183 & 57 & 223 \\ 91 & -132 & -30 & -58 & -129 \end{array}\right],$$

we get the following sequence of row vectors by multiplying on the right by A_2 and orthogonalizing with respect to X_1 at each step:

$$\{0, 1, 0, 0\},$$
$$\{1.00, -1.56, -.46, -0.57\},$$
$$\{1.0000, -1.5795, 0.26538, -0.5526\},$$
$$\{1.000000, -1.579690, 0.2653589, -0.552692\},$$
$$\{1.0000000, -1.5796883, 0.2653587, -0.5526927\}.$$

Operating once with A yields $\lambda_2 = -1.7684837$. It follows that $\lambda_{3,4} = -0.3591939 \pm 3.2840604i$.

How to find the characteristic vectors corresponding to such complex roots will be explained below.

21. Complex roots. When the complex roots cannot be approached from behind in this fashion, a new method of attack is necessary. Since such roots come in conjugate pairs of equal modulus, applying the matrix repeatedly to an arbitrary vector will not result in a convergent vector sequence (in fact such behaviour is sometimes the best indication that the largest roots are complex).

However, if $\lambda_1 = r_1 e^{i\theta_1}$, $\lambda_2 = r_1 e^{-i\theta_1}$, $|\lambda_1| = |\lambda_2| > |\lambda_3| > \cdots$, we can still find the product $\lambda_1\lambda_2 = r_1^2$ by the method used in Example 2; indeed the convergence should be relatively better than when $|\lambda_1| > |\lambda_2|$, since the loss of accuracy due to subtracting almost equal numbers will not arise. Having r_1^2, we can find λ_1 and λ_2 as follows.

First by operating repeatedly on a vector V with A (or better, with some power of A) we obtain a vector W of the form

$$W = a_1X_1 + a_2X_2 + \Delta, \tag{38}$$

where Δ is as small as we please. It can now be seen that

$$r_1 \cos\theta_1 = \frac{r^2W^{(j)} + (A^2W)^{(j)}}{2(AW)^{(j)}} + 0(\Delta) \qquad (j = 1, 2, 3, \cdots, n). \tag{39}$$

Thus λ_1 *and* λ_2 *are determined to any desired accuracy.*

Furthermore,

$$\begin{aligned} AW + i(r^2W - A^2W)/2r_1 \sin\theta_1 &= a_1\lambda_1X_1 + a_2\lambda_2X_2 + i\frac{\lambda_1\lambda_2(a_1X_1 + a_2X_2) - a_1\lambda_1^2X_1 - a_2\lambda_2^2X_2}{i(\lambda_1 - \lambda_2)} + 0(\Delta) \\ &= 2a_1\lambda_1X_1 + 0(\Delta) \end{aligned} \tag{40}$$

so that X_1 *and* X_2 *can be found once* λ_1 *and* λ_2 *are known.*

The problem can also be approached by operating with a matrix of the form $A + ni$, whose roots will ordinarily have distinct moduli, but in practice this method will usually prove more laborious than the preceding one

If all the roots are wanted, we begin by locating them roughly, as in the symmetric case. Then matrix polynomials can be constructed to secure rapid convergence to more accurate values. These points will become clearer in the following example.

Example 8.

$$A = \left[\begin{array}{cccc|c} 1 & -2 & 0 & -4 & -5 \\ 3 & 0 & 1 & 2 & 6 \\ -1 & 3 & -1 & 1 & 2 \\ 1 & 0 & 4 & 0 & 5 \end{array}\right].$$

Find the roots of A to seven decimal places and its characteristic column vectors to seven significant figures.

As usual, we write down A^2 and A^3:

$$A^2 = \left[\begin{array}{cccc|c} -9 & -2 & -18 & -8 & -37 \\ 4 & -3 & 7 & -11 & -3 \\ 10 & -1 & 8 & 9 & 26 \\ -3 & 10 & -4 & 0 & 3 \end{array}\right] \qquad A^3 = \left[\begin{array}{cccc|c} -5 & -36 & -16 & 14 & -43 \\ -23 & 13 & -54 & -15 & -79 \\ 8 & 4 & 27 & -34 & 5 \\ 31 & -6 & 14 & 28 & 67 \end{array}\right].$$

Thus $\sum_{i=1}^{4} \lambda_1 = 0$, $\sum_{i=1}^{4} \lambda_i^2 = -4$, and the largest root must be complex.

Starting with the vector $V = (0, 0, 1, 0)$, we operate repeatedly with A^3, rounding off to three or four figures at each step to avoid accumulating useless digits. After several such multiplications we come to the sequence of vectors given below:

$$(-458, 752, -68, -432), (-2974, 3046, 1220, -3176),$$
$$(1588, 898, 1293, -1823), (-7060, 572, 8778, -8756).$$

As we proceed, we form two-rowed determinants from the first and second elements of successive vectors, and likewise from the second and third elements, and then take the ratios of the values of successive determinants. The terms of tehse sequences corresponding to the above vectors are

$$\begin{array}{ccccc} 841380 & & 2166396 & & 5431544 \\ & 257 & & 251 & \\ 1224568 & & 2482918 & & 7143048 \\ & 253 & & 251 & \end{array}$$

(The sequences of ratios above show convergence to r_1^6, aside from the location of the decimal point, which is immaterial at this stage.)

The agreement of the last numbers suggest that if we now operate on the last vector with A and form ratios of determinants as before, we shall obtain r_1^2 to about three figures:

$$(-7060, 572, 8778, -8756), (26820, -29914, -8758, 28052),$$
$$(-25560, 127806, -79752, -8212), (-248324, -172856, 480518, -344568),$$

$$\begin{array}{ccccc} 19585 & & 266316 & & 3615550 \\ & 13.60 & & 13.58 & \\ 25758 & & 350503 & & 4762747 \\ & 13.61 & & 13.59 & \end{array}$$

Thus $r_1^2 \sim 13.6$. Using equation (39) with W as the second vector above yields

$$r_1 \cos\theta_1 \sim -2.27, \qquad r_1 \sin\theta_1 \sim 2.91,$$
$$\lambda_1 \sim -2.27 + 2.91i, \qquad \lambda_1 \sim -2.27 - 2.91i.$$

It follows that $\lambda_{3,4} = 2.27 \pm 1.96i$.

In order to get the roots more exactly, we note that

$$(\lambda - 2.27 + 1.96i)(\lambda - 2.27 - 1.96i) = \lambda^2 - 4.54\lambda + 8.99.$$

Thus the matrix $A_1 = 2A^2 - 9A + 18$ will have two roots near zero. Rounding off the elements of the last vector above, we multiply it by A_1 and obtain ratios of determinants formed from the first and second, and second and fourth elements:

$$(-2483, -1729, 4805, -3446)(-243759, 188294, 153735, -27083),$$
$$(-6120173, 18490944, -8355979, -4216169),$$
$$(53044664, 46504148, -111097501, 75339351),$$
$$(56799532, -40188604, -39094634, 63788114),$$

$$\frac{-888993313}{3774} \quad \frac{-335494216}{37719260} \quad \frac{-126545934}{37719148} \quad \frac{-477320484}{37719148}$$

$$\frac{111704493}{3771} \quad \frac{421315396}{37719131} \quad \frac{158916507}{37719148} \quad \frac{599419524}{37719148}$$

We now multiply the last vector above by A and form ratios of determinants, getting

$$(56799532, -40188604, -39094634, 63788114),$$
$$(-117975716, 258880190, -74482596, -99579004),$$
$$(-237420080, -627567752, 869519878, -415906100),$$

$$\frac{9962994304}{13.6004404} \quad \frac{1355011183}{13.6004405}$$

$$\frac{1311417783}{13.6004405} \quad \frac{1783585959}{13.6004405}$$

Taking $r_1^2 = 13.6004405$, we obtain, by using (39), $\lambda_{1,2} = -2.26774878 \pm 2.90822213i$.

Using equation (40) with W as the first vector above gives

$$2a_1X_1 = (-117975716 + 173631636i, 258880190 + 13923461i,$$
$$-74482596 - 240907341i, -00579004 + 220659305i)$$
$$X_1 = (1.00000000, -0.63822188 - 1.05732751i,$$
$$-0.74982611 + 0.93844573i, 1.13604812 - 0.19839177i).$$

The remaining roots must be $\lambda_{3,4} = 2.26774878 \pm 1.95642866i$.

22. Finding the remaining vectors. The vector X_3 cannot be determined as X_1 was, since there is no ready-made vector sequence available. However, the matrix polynomial

$$A_2 = (A + 2.26774878 - 2.90822213i)(A + 2.26774878 + 2.90822213i)$$
$$\cdot(A - 2.26774878 - 1.95642866i)$$
$$= (A^3 + 2.26774878A^2 + 3.31507144A - 30.8423824)$$
$$+ i(1.95624866A^2 + 8.87337741A + 26.6082916)$$

has three roots so near zero that a single multiplication of an arbitrary vector by A^2 should give us X_3 to about eight figures. To do this *it is not necessary to compute the elements of the matrix* A_2, for by (9)

$$A_2V = A^3V + 2.26774878A^2V + 3031507144AV - 30.8423824V + i(1.95642866A^2V + 8.87337741AV + 26.6082916V).$$

Taking

$$V\,(1, 1, 1, 1),$$

we see that AV, A^2V, A^3V are the sum solumns of A, A^2, A^3. Then

$$A_2V = (-174.3244445 - 90.1464559i, -96.7552001 + 73.9792701i, 39.7292288 + 95.2221916i, 59.5362211 + 76.8444646i)$$

$$X_3 \sim (1.00000000, .26477276 - .56129590i, -.40277933 - .33795067i, -.44932357 - .20845922i).$$

(The second vector above was obtained from the first by dividing by the leading element.) As a check, we perform the same operations with

$$V_1 = (1, 0, 0, 0)$$

(note that AV_1, A^2V, A^3V_1 are simply the first columns of A, A^2, A^3) and get

(1.00000000, .26477273 −.56129590i, −.40277934 −.33795066i, −.44932357 −.20845922i).

VI. Solution of Algebraic Equations by Matrix Methods

23. The companion matrix. Before closing we consider briefly the solution of an algebraic equation

$$f(\lambda) = \lambda^n - p_1\lambda^{n-1} - p_2\lambda^{n-2} - \cdots - p_n = 0 \tag{41}$$

by the use of the "companion matrix"

$$C = \begin{bmatrix} p_1 & p_2 \cdots & p_{n-1} & p_n \\ 1 & 0 \cdots & 0 & 0 \\ 0 & 1 \cdots & 0 & 0 \\ \cdot & \cdot \cdot \cdot & \cdot & \cdot \\ 0 & 0 \cdots & 1 & 0 \end{bmatrix} \tag{42}$$

(Bernouilli's method). It is easily seen that (42) has the same roots as (41) and that the characteristic column vector of (42) corresponding to the root $\lambda_i (i = 1, 2, \cdots, n)$ is

$$X_i = (\lambda_i^{n-1}, \cdots, \lambda_i^2, \lambda_i, 1) \quad \text{or} \quad (1, \lambda_i^{-1}, \lambda_i^{-2}, \cdots, \lambda_i^{-n+1}). \tag{43}$$

Thus the roots of (41) are determined if either the roots or the column vectors of (42) can be found.

24. Special methods of solution. While the methods employed in handling a gen-

eral matrix are applicable in this case, the special properties of C can be used to speed up the solution. First, equation (41) helps us in locating the roots, especially the real ones. Again, if λ_i' is an approximation of λ_i, the matrix polynomial

$$\frac{f(C) - f(\lambda_i')}{\lambda - \lambda_i'} \tag{44}$$

(which may be obtained by synthetic division) will have only one large root, corresponding to λ_i. We recall that to set up the corresponding polynomial $P(A)$ in the general case approximations to all the roots are required. (However, (44) is not easily set up if λ_i is complex.)

Moreover, suppose we have a vector V that is an approximation of X_i and that $D=P(C)$ is a matrix having $P(\lambda_i)$ as its largest root. Then our next approximation, under our previous procedure, would be

$$W = (1, w_2, w_3, \cdots, w_n) = \frac{DV}{(DV)^{(1)}}. \tag{45}$$

Instead we may use the more easily computed

$$W^* = (1, w_2, w_2^2, w_2^{n-1}).$$

This can be done because X_i and W^* are both of the form (43), and because w_2, being obtained from W, is an approximation to λ_i^{-1}. Unfortunately this process is not necessarily convergent (though in practice it generally is, especially when $|P(\lambda_i)| \gg |P(\lambda_i)|$, $j \neq i$). However, we can always try it and then fall back on the standard process if necessary.

Clearly, rather than determining W^* from the first two elements of DV, we could determine it from the last two. It can be shown that this will lead to better convergence if λ_i is the smallest root of (41), while the first process will work better if λ_i is the largest root of (42). For roots in between we should use one process or the other depending on their relative magnitude.

25. Examples. The following examples will make this procedure clear.

Example 9.

$$f(\lambda) = \lambda^4 - 4\lambda - 3 = 0. \tag{46}$$

Direct substitution shows that (46) has real roots between -1 and 0 and between 1 and 2.

If we take 2 as a first approximation to λ_i, the matrix corresponding to (44) is $C_1 = C^3+2C^2+4C+4$.

The vector corresponding to λ_1 can be written

$$X_1 = (1, \lambda_1^{-1}, \lambda_1^{-2}, \lambda_1^{-3}).$$

The first element is taken as unity since λ_1 is a relatively large root. Since $\lambda_1 \sim 2$, we start with the vector $V=(1.00, .50, .25, .12)$. Operating with C_1 yields the vectors

$$(1.000, .561, .315, .177), \qquad (1.0000, .5604, .3140, .1760).$$

In forming these vectors, the first two elements of each vector were obtained in

the usual way, but the last two elements are simply the square and cube of the second element. Note that only the first two rows of C_1 are used in this procedure.

From the last vector we get

$$\lambda_1 \sim \frac{1}{0.5604} = 1.794.$$

In order to get λ_1 more exactly, we divide (46) by $\lambda - 1.784$ getting approximately

$$\lambda^3 + 1.784\lambda^2 + 3.18\lambda + 1.68.$$

Thus the matrix $C_2 = 5C^3 + 9C^2 + 16C + 8$ has only one large root, corresponding to λ_1 (cf. (44)). Operating as before (using only the first two rows of C_2) gives the rapidly convergent vector sequence

$$(1.000, .560, .314, .176)(1.00000, .56043, .31408, .17602),$$

$$(1.0000000, .5604256, .3410769, .1760167),$$

$$(1.00000000, .56042566, .31407692, .17601677),$$

whence $\lambda_1 = 1.7843580$.

Note that it was not necessary to approximate the other roots in order to find λ_1.

The root λ_2 can be obtained in much the same way. Starting with the matrix

$$C_3 = C^3 - C^2 + C - 5$$

and the vector

$$V = (-1, 1, -1, 1)$$

we reach after a few steps the vector

$$(-.334, .482, -.694, 1.000).$$

(Since λ_2 is the smallest root, we take the last element as unity here and use the last two rows of C_3 in the computation.) Thus $\lambda_2 \sim -0.69$. Dividing (36) by $\lambda + 0.69$ gives

$$\lambda^3 - 0.69\lambda^2 + 0.48\lambda - 4.33.$$

Using the matrix

$$C_4 = -10C^3 + 7C^2 - 5C + 43$$

yields after a few more steps $\lambda_2 = -0.69250484$.

The complex roots are then $\lambda_{3,4} = -0.5459266 \pm 1.4593779i$.

Example 10.

$$\lambda^4 - 4\lambda + 4 = 0. \tag{37}$$

Here the roots are all complex. The companion matrix is

$$D = \left[\begin{array}{cccc|c} 0 & 0 & 4 & -4 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \end{array}\right].$$

By the methods employed in the general case, we get the rough values $\lambda_{1,2} \sim -1.05 \pm 1.43i$, $\lambda_{3,4} \sim 1.05 \pm 0.41i$.

Noting that

$$(\lambda - 1.05 - .41i)(\lambda - 1.05 + .41i)(\lambda + 1.05 + 1.43i)$$
$$= (\lambda^3 - 1.05\lambda^2 - .94\lambda + 1.33) + i(1.43\lambda^2 - 3.00\lambda + 1.82),$$

we set up the matrix

$$D_1 = (10D^3 - 11D^2 - 9D + 13) + i(14D^2 - 30D + 18)$$
$$= \begin{bmatrix} 53 & -84 & 8 & 36 \\ -9 & 53 & -4 & 44 \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix} + i \begin{bmatrix} 18 & 56 & -176 & 120 \\ -30 & 18 & 36 & -56 \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix}.$$

Since only the first two rows will be used, the others need not be written. Starting with a vector whose elements are power of $(-1.05+1.43i)^{-1}$, we get the vector sequence

$(1.000, -.333 - .454i, -.095 + .302i, .169 - .057i)$,

$(1.000000, -.332486 - .453252i, -.094890 + .301400i, .168160 - .057202i)$,

$(1.00000000, -.33248603 - .45325413i, \text{————}, \text{————})$,

Thus

$$\lambda_1 = \frac{1}{-0.33248603 - 0.45325413i} = -1.0522167 + 1.4344109i$$

and $\lambda_{3,4} = 1.0522167 \pm 0.3959609i$.

THE TURBULENT FLOW ALONG A SEMI-INFINITE PLATE*

BY

P. Y. CHOU

National Tsing Hua University, Peiping, China

1. Introduction. The investigation of the turbulent flow along a semi-infinite plate is of both theoretical and technical interest, as it leads to the magnitude of the skin-friction acting upon the plate at high Reynolds numbers. Interpretations of the velocity distribution within the tubulent boundary layer have been given by Howarth[1] according to the transport theories. But the formula that fits experiment best is von Kármán's logarithmic law of velocity distribution which is a consequence of his similarity theory of turbulence.

The purpose of the present paper is to show that von Kármán's velocity distribution law is consistent with the program of treating turbulent flows developed by the author.[2] For the solution of the present problem we shall adopt the following three conditions and simplifications.

(1) We shall use the law of dynamical similarity down-stream along the flow. which has been verified by Dryden's experiments. The mean velocity along the direction of the undisturbed flow U which is parallel to the plane of the plate when expressed in units of the friction velocity U_τ has been found by Dryden to be a function of the dimensionless variable $\eta = yU_\tau/\nu$ in which y is the coordinate perpendicular to the plate and U_τ is a function of x, the distance along the plate from its leading edge. We now generalize this result by assuming that the mean pressure $\bar{p}$ and all the other velocity correlations when expressed in units of proper powers of U_τ are dimensionless functions of η only.

(2) Reynolds' equation for mean motion and the equations of double and triple correlations will be used. In the equations of triple correlation we shall disregard the quadruple correlations. The range of validity of this approximation and the possibility of its experimental verification will be discussed in Sec. 5.

(3) We shall assume the constancy of the micro-scale of turbulence λ and of the squares of components of velocity fluctuation across the boundary layer. This condition is used for mathematical simplicity. It is, however, consistent with the result to be shown below that the shearing stress is constant across the boundary layer and can therefore be regarded as a condition of similarity. On the other hand, as will be seen presently, the variation of the mean squares of velocity fluctuation can be accounted for, though their solution will be rather complicated. This part of calculation is omitted due to lack of measurements.

The logarithmic law of temperature distribution in the boundary layer of a heated plate can also be established by the present method of approach to the turbulence problem and its solution will be indicated at the end of the paper.

2. Reynold's equations of mean motion. Let the positive half of the xz-plane represent the given semi-infinite plane and the leading edge of the plane coincide with the z-axis, the y-axis being perpendicular to the plane. Using U and V to denote the

* Received Oct. 21, 1946.

[1] S. Goldstein, *Modern developments in fluid dynamics* (1938) pp. 361–371.

[2] P. Y. Chou, Quart. Appl. Math. **3**, 38–54 (1945).

x- and y-components of the mean velocity, we find the Reynolds' equations of mean motion to be

$$U\frac{\partial U}{\partial x}+V\frac{\partial U}{\partial y}=\frac{1}{\rho}\frac{\partial \tau_{xy}}{\partial y}, \tag{2.1}$$

$$U\frac{\partial V}{\partial x}+V\frac{\partial V}{\partial y}=-\frac{1}{\rho}\frac{\partial \bar{p}}{\partial y}+\frac{1}{\rho}\frac{\partial \tau_{yy}}{\partial y}, \tag{2.2}$$

in which both the viscous stress and the terms in components of turbulent stress involving partial differentiations with respect to x on the right hand sides of Eqs. (2.1) and (2.2) are neglected. This is permissible on account of the thinness of the turbulent boundary layer.

To satisfy the equation of continuity, we introduce a stream function Ψ as usual and write

$$U=\frac{\partial \Psi}{\partial y}, \qquad V=-\frac{\partial \Psi}{\partial x}. \tag{2.3}$$

It is well-known that by integrating Eq. (2.1) across the boundary layer, we find the turbulent shearing stress on the plate. Thus,[3]

$$\tau_0=\tau_{xy}\big|_{y=0}+\rho U_\tau^2=\frac{\partial}{\partial x}\int_0^\delta \rho(U_1-U)U\,dy, \tag{2.4}$$

where U_τ is the friction velocity, δ the thickness of the layer and U_1 the velocity in the free stream outside the layer. Hence if the mean velocity U, as a function of y, were known, Eq. (2.4) would define the friction velocity U_τ as a function of x.

Here we introduce Dryden's experimental result that the ratio of the mean velocity U and U_τ is a function of the dimensionless variable η, where η is defined by

$$\eta=yU_\tau/\nu. \tag{2.5}$$

In other words, the stream function Ψ is given by

$$\Psi=\nu F(\eta) \tag{2.6}$$

so that according to equation (2.3)

$$U=U_\tau F'(\eta). \tag{2.7}$$

This relation can be construed as the condition of dynamical similarity in the different cross-sections of the boundary layer. From Eqs. (2.3) and (2.6), we obtain

$$V=-yF'\frac{\partial U_\tau}{\partial x}. \tag{2.8}$$

Let us assume that dynamical similarity also holds true for the shearing stress, i.e. that

$$-\frac{1}{\rho}\tau_{xy}=U_\tau^2 f_4(\eta). \tag{2.9}$$

[3] S. Goldstein, *loc. cit.*, p. 132.

Substituting from Eqs. (2.7), (2.8) and (2.9) into the equation of mean motion (2.1), we obtain

$$F'^2 \frac{\partial U_\tau}{\partial x} = -\frac{U_\tau^2}{\nu} f_4'. \tag{2.10}$$

We may estimate the order of magnitude of $\partial U_\tau/\partial x$ by means of von Kármán's coefficient of local skin friction in the following way:[1]

$$\frac{dz}{dx} = \frac{1}{x}\frac{1}{1+2/z}, \tag{2.11}$$

where $z = KU_1/U_\tau$, $K = 0.4$.

The derivative dz/dx thus tends to $1/x$ for large values of z, Hence equation (2.10) becomes

$$f_4' = \frac{F'^2}{K}\frac{\nu}{U_1 x} = F'^2/KR_x. \tag{2.12}$$

The Reynolds number R_x in Dryden's measurements[4] is of the order of 10^6 and F' is around 20. This shows that f_4' is of the order 10^{-3} and can be neglected. In short, f_4 or the shearing stress is essentially constant across the turbulent boundary layer.

3. Equations of double and triple correlations, the mean velocity distribution. We suppose that the law of dynamical similarity not only holds true for the mean velocity distribution, but also for the double and triple velocity correlations as well. Hence we write

$$\left.\begin{aligned} -\frac{1}{\rho}\bar{p} &= U_\tau^2 f_0(\eta), \\ -\frac{1}{\rho}\tau_{xx} = U_\tau^2 f_1(\eta), \quad -\frac{1}{\rho}\tau_{yy} &= U_\tau^2 f_2(\eta), \quad -\frac{1}{\rho}\tau_{zz} = U_\tau^2 f_3(\eta), \\ -\frac{1}{\rho}\tau_{xy} &= U_\tau^2 f_4(\eta); \end{aligned}\right\} \tag{3.1}$$

$$\left.\begin{aligned} \overline{W_x^3} &= U_\tau^3 h_0(\eta), \quad \overline{W_x^2 W_y} = U_\tau^3 h_1(\eta), \\ \overline{W_x W_y^2} &= U_\tau^3 h_2(\eta), \quad \overline{W_y^3} = U_\tau^3 h_3(\eta), \\ \overline{W_x W_z^2} &= U_\tau^3 h_4(\eta), \quad \overline{W_y W_z^2} = U_\tau^3 h_5(\eta). \end{aligned}\right\} \tag{3.2}$$

Since our interest at present lies mainly in the mean velocity distribution, we shall only put down the equations which are pertinent for its determination. When written in full the component equation ($i=1$, $k=2$) for the double correlation[2] is

$$-\frac{1}{\rho}\left[\tau_{xx}\frac{\partial V}{\partial x} + \tau_{yy}\frac{\partial U}{\partial y}\right] - \frac{1}{\rho}U\frac{\partial \tau_{xy}}{\partial x} - \frac{1}{\rho}V\frac{\partial \tau_{xy}}{\partial y}$$
$$+\frac{\partial}{\partial x}\overline{w_x^2 w_y} + \frac{\partial}{\partial y}\overline{w_x w_y^2} = -a_{2112}\frac{\partial U}{\partial y} - \frac{2\nu k}{\lambda^2}\overline{w_x w_y}. \tag{3.3}$$

[4] H. L. Dryden, N. A. C. A. Report No. 562 (1936).

In the above equation we have only retained the largest term $a_{2112}\partial U/\partial y$ among the sum $a^n_{mik}U^m{}_{,n}$.

Next, introduce the substitutions (2.7), (2.8), (3.1) and (3.2) into Eq. (3.3). Obviously, the terms involving $\partial V/\partial x$ and $\overline{w_x^2 w_y}$ in (3.3) are small. With the use of the relation (2.11), we find that

$$-\frac{1}{\rho}U\frac{\partial \tau_{xy}}{\partial x}-\frac{1}{\rho}V\frac{\partial \tau_{xy}}{\partial y}=-\frac{2U_\tau^4}{KxU_1}F'f_4. \tag{3.4}$$

The appearance of xU_1 in the denominator means that this term contains the reciprocal of the Reynolds number R_x as a factor when compared with the remaining terms, says, $-\tau_{yy}\partial U/\rho\partial y$, in the equation. Hence as a good approximation of (3.3) we have

$$-aF''+h_2'=-l_0f_4, \tag{3.5}$$

where l_0 and a are defined by

$$l_0=2k\nu^2/\lambda^2U_\tau^2, \qquad a=-(f_2+a_{2112}/U_\tau^2). \tag{3.6}$$

The fact that only the terms in $\partial U/\partial y$ and $\partial\overline{w_x w_y^2}/\partial y$ on the left-hand side of equation (3.3) remain can be interpreted that V is negligible and all the functions U, τ_{ij} and $\overline{w_i w_j w_k}$ depend upon η alone and are very approximately independent of x, the coordinate from the leading edge of the plate. An equivalent statement is that the turblent boundary layer is very thin. This simplifies the problem; its mathematics becomes analogous to that for the problem of mean velocity distribution in a channel.[5]

The component equation ($i=1$, $k=1$, $1=2$) of the equations of triple correlation which is needed for the determination of the mean velocity distribution can be obtained by the method used in deriving equation (3.5) as:

$$2\frac{\partial U}{\partial y}\overline{w_x w_y^2}+\frac{\partial}{\partial y}\overline{w_x^2 w_y^2}=-b_{21112}\frac{\partial U}{\partial y}-c_{112}+\frac{1}{\rho^2}\left(2\tau_{xy}\frac{\partial \tau_{xy}}{\partial y}+\tau_{xx}\frac{\partial \tau_{yy}}{\partial y}\right). \tag{3.7}$$

It is obvious that the quadruple correlation term $\partial\overline{w_x w_y^2}/\partial y$ must be of the same order of magnitude as $\tau_{xx}\partial\tau_{yy}/\rho^2\partial y$ and we shall now neglect these quantities against the term $\partial U/\partial y(2\overline{w_x w_y^2}+b_{21112})$. This is equivalent to the statement that we limit our solution to the region where the following inequality holds true

$$\left|\frac{\partial U}{\partial y}(2\overline{w_x w_y^2}+b_{21112})\right|>\left|\overline{w_x^2}\frac{\overline{\partial w_y^2}}{\partial y}-\frac{\partial}{\partial y}\overline{w_x^2 w_y^2}\right|. \tag{3.8}$$

Apparently this condition can be satisfied in the middle of the layer where $\partial U/\partial y$ is large, while on its outer edge it may break down as we shall see in Sec. 5. The quantity c_{112} in (3.7) will be considered as independent of η because it varies but slowly with the coordinates, as has been proved previously.[2] The derivative $\partial\tau_{xy}/\partial y$ is small according to the relation (2.12) and can therefore be neglected. Introducing expressions (2.7) and (3.2) into Eq. (3.7) and writing

$$c_{112}=2cU_\tau^5/\nu,$$

[5] P. Y. Chou, Quart. of Appl. Math., **3**, 198–209 (1945).

we find, after using the inequality (3.8),

$$h_2 = -c/F'' - b_{21112}/2U_\tau^3. \tag{3.9}$$

Like c_{112} we take b_{21112}/U_τ^3 to be also a constant.

Elimnation of the variable h_2 between equations (3.5) and (3.9) yields the final equation for the determination of means velocity distribution,

$$aF'' + c\frac{d}{d\eta}\frac{1}{F''} = l_0 f_4. \tag{3.10}$$

Since the variation of f_4 with η across the boundary layer is small, we assume that a which depends upon $\overline{w_y^2}$ defined by equation (3.6) is also a constant. There is no *a priori* experimental justification for the assumption, but it simplifies the calculation a great deal. As a matter of fact the determination of the behaviour of $\overline{w_x^2}$, $\overline{w_y^2}$ and $\overline{w_z^2}$ within the layer is not wholly impossible by means of the rest of the equations of double and triple correlations which have not been written out explicitly,* though the work involved will be rather complicated. For the same reason of simplifying calculation we take l_0 to be constant, namely, λ to be independent of η and leave out the equation of vorticity decay.[2]

Integrating the equation (3.10) twice, we find

$$\eta = \frac{a}{l_0 f_4}F' + A(e^{l_0 f_4 F'/c} - 1), \tag{3.11}$$

where A is a constant of integration. The other arbitrary constant has been determined by the condition on the surface of the plate: $\eta = 0$, $F' = 0$. Another constant of the remaining three in equation (3.11) can be disposed of, as a consequence of the linear dependence of the mean velocity U upon η near the plate, by requiring that it become $\eta = F'$, when η is small. On the other hand for large values of η the mean velocity U varies with η logarithmically, confirming von Kármán's result. For this latter reason we take the factor of F' in the exponential function to be 0.4, when we compare formula (3.11) with Dryden's measurements.[4] The last constant A is determined by passing the theoretical curve through the mean experimental value of $F' = 18.2$ at $\log_{10}\eta = 2.2$. The result can be summarized in the equation

$$\eta = 0.959\frac{U}{U_\tau} + 0.1024(e^{0.4U/U_\tau} - 1). \tag{3.12}$$

The velocity distribution given above resembles that within the channel[5] in its functional dependence upon η, though the values of the corresponding constants appearing in the two cases are different as is to be expected. Comparison between theory and experiment is given in Fig. 1. Their agreement is well-known to be satisfactory.

---

* They will be the same as Eqs. (2.3), (2.5), (2.6), (3.1), (3.3) and (3.4) in the paper quoted in Footnote 5.

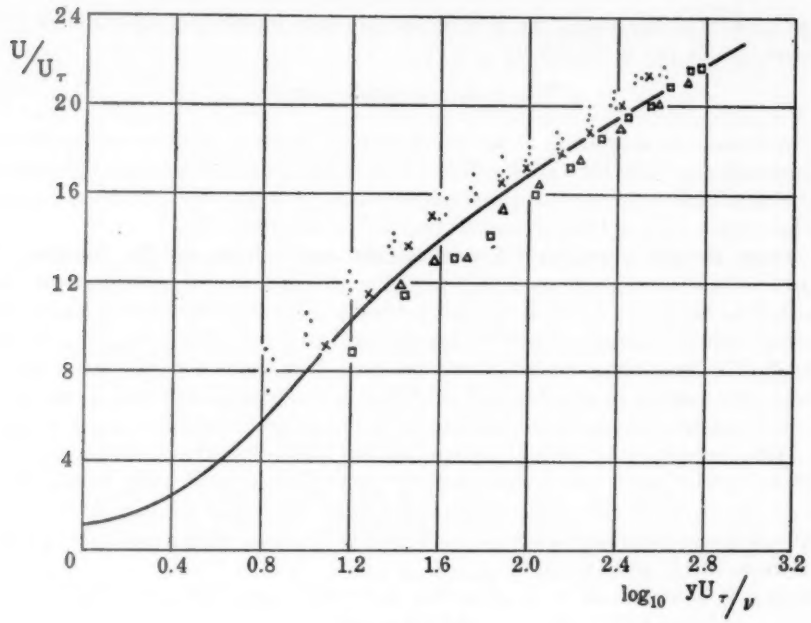

FIG. 1

Velocity distribution in a turbulent boundary layer.
Theory: ————————
Dryden's experimental results:
· $R_x = 335{,}000$, □ $R_x = 507{,}000$,
× $R_x = 788{,}000$, △ $R = 983{,}000$,
x being measured from the transition point.

4. The coefficient of local skin-friction. The explicit relation between the coefficient of local skin-friction c_f, which is defined by

$$c_f = \tau_0/\tfrac{1}{2}\rho U_1^2 = U_\tau^2/\tfrac{1}{2}U_1^2, \tag{4.1}$$

and the Reynolds number $R_x = U_1 x/\nu$ can now be determined by substituting the mean velocity distribution (3.12) into Eq. (2.4) and integrating. The thickness of the boundary layer $y = \delta$ is found by setting $U = U_1$ in Eq. (3.12). Retaining the large quantities in the final result after integration, we find

$$\frac{dR_x}{dz} = 1.6e^z(z^2 - 2z + 2), \tag{4.2}$$

where z has been defined in Eq. (2.11) as $0.4U_1/U_\tau$. Integrating (4.2) again gives

$$R_x = 1.6e^z(z^2 - 4z + 6) - 6. \tag{4.3}$$

The constant of integration is determined by $x = 0$, when $z = 0$, under the assumption that the flow is turbulent from the leading edge.

For large z we keep only the z^2 term on the right member of equation (4.3) and the result can be put in the form

$$c_f^{-1/2} = 1.19 + 4.08 \log_{10} (R_x c_f). \tag{4.4}$$

The two constants appearing above according to Kempf's experiments are 1.7 and 4.15 respectively.[1] The form of R_x given in (4.3) also justifies our use of the relation (2.11) which leads to the very approximate constancy of the shearing stress across the boundary layer as has been demonstrated by relation (2.12).

5. Some further experimental verifications and validity of the solution. The present analysis shows that, according to (2.12), the shearing stress is constant across the turbulent boundary layer, if the Reynolds number of the flow is sufficiently high. This result can be subject to direct experimental proof, since τ_{xy} at different points of the fluid is now measurable by the hot-wire technique. On the other hand to establish the validity of the inequality (3.8) experimentally is probably not simple. There is a greater chance of its fulfillment, if $\overline{w_x^2}$ and $\overline{w_y^2}$ are also constant across the layer. This constancy of $\overline{w_x^2}$ and $\overline{w_y^2}$ can be interpreted as the consequence of the condition of "similarity" mentioned before. On the other hand, even if this requirement of similarity were satisfied, it has its limitations and we can determine the validity of the present solution by considering the values of the constants appearing in equations (3.11) and (3.12).

Since f_4 is independent of η, according to its definition (2.9) and relation (2.4) it must be -1. From Eqs. (3.11) and (3.12) we find

$$a/l_0 = -0.959, \qquad l_0/c = -0.4. \tag{5.1}$$

The quantity l_0 is determinable by using Eq. (3.6) with the micro-scale of turbulence λ, which comes from correlation measurements and k, a constant, which is present in all the three equations for $\overline{w_x^2}$, $\overline{w_y^2}$, and $\overline{w_z^2}$ and is also observable. Thus both a and c are known after l_0 is fixed.

The factor $1/F''$ in equation (3.9) for h_2 varies according to (3.12) between unity, when $U/U_\tau = F' = 0$ on the plate and about 120 when $F' = 20$. Hence the part of h_2 variable with η can take the extreme values of $-c$ and $-120c$. According to its definition in (3.2), the value of h_2 can not be very large. This means that there must be an upper limit of F', namely, the outer limit of the boundary layer within which the relation (3.9) is valid. Again if triple velocity correlations were measurable by the hot wire technique, to test this relation directly within its region of validity should be an interesting problem.

The foregoing analysis shows the important rôle played by the triple velocity correlation within the turbulent boundary layer. This can be interpreted as indicating that turbulence is generated in this layer and transported away from it. The triple velocity correlation is a measure of this transport; hence the set of equations governing its behaviour is needed for the interpretation of the mean velocity distribution. Indeed, if this interpretation were correct, it is easy to understand that only the equations of mean motion and of double correlation with the omission of the triple correlation terms in the latter are insufficient to explain the velocity distribution within the boundary layer, for this simplified set of the equations of double correlation only tells us the balance between the production of turbulent energy by deformation of the mean flow and the decay of turbulence. Although this picture

seems to give a good account of the velocity distributions in the center of the channel and in free turbulence like jets and wakes, it is definitely insufficient for the phenomena dealt with in the present paper.

The present paper is obviously incomplete in the sense that theoretical computations of the mean squares of turbulent fluctuations are not given. It does show, however, that a certain type of function for the mean velocity distribution is necessary. Though refinements of calculation and measurements may change this result to some extent, the main feature of the function will probably remain the same. This point has been borne out by the corresponding calculation of the velocity distribution in the channel under the assumption of different turbulent levels.[5]

The temperature distribution within the boundary layer of a heated plate can be calculated by means of the equations of mean temperature and of the double and triple velocity and temperature correlations. We obtain an equation for the mean temperature distribution analogous to (3.11), and it agrees with the measurements of Éliás.[6] Obviously a limitation on the validity of the solution similar to (3.8) is also present.

[6] F. Éliás, Zeits. angew. Math. Mech., **10**, 1–14 (1930).

—NOTES—

A CONTRIBUTION TO THE METHOD OF LEAST SQUARES*

By M. HERZBERGER and R. H. MORRIS (*Eastman Kodak Company*)

In trying to calculate the image figures of an optical problem the authors were led to consider a problem of least squares which proved unusually difficult, since the determinant of the normal equations to be solved had a very small value. The following attempt to make the problem a geometrical one should be of value in other applications, and is presented here for what it is worth.

Most problems of least squares can be reduced to the following: Suppose that we are given a sequence of k known functions $f_1, \cdots, f_k$ of one or more variables. Our problem is to determine k constants, C_i, so that an unknown function, F, whose values are known for $n > k$ values of the coordinates will be approximated as closely as possible by a linear combination of the functions f_i:

$$F = C_1 f_1 + \cdots + C_k f_k. \tag{1}$$

Since F is given for only $n > k$ values, we can consider these values as the coordinates of a vector $\bar{b}$ in an n-dimensional space. The values of the f_i for the n values of the coordinates may also be considered as the coordinates of k vectors $\bar{a}_i$ in the n-dimensional space. It is a simple geometrical problem to find the best approximation of the vector $\bar{b}$ by a linear combination of the vectors $\bar{a}_i$. We merely project $\bar{b}$ in to the space of the vectors $\bar{a}_i$, i.e., we minimize

$$\left(\bar{b} - \sum_{i=1}^{k} C_i \bar{a}_i\right)^2. \tag{2}$$

This leads to the k normal equations for the constant C_i

$$\sum_i C_i a_i d_\lambda = b a_\lambda, \tag{3}$$

which must be solved.

We can simplify the problem considerably if we order the vectors in a suitable way, which in this case will be the transformation of (3) into a form which contains only unit vectors, i.e., the projection of our problem on the unit sphere. Let

$$\bar{a}_i = \alpha_i \bar{s}_i, \qquad \bar{b} = \beta \bar{t}, \tag{4}$$

where $s_i^2 = t^2 = 1$. Equation (3) becomes

$$\sum \gamma_i s_i s_\lambda = t s_\lambda. \tag{5}$$

A solution of (5) which yields γ_i will also yield C_i for

$$C_i = \frac{\beta \gamma_i}{\alpha_i}, \qquad \gamma_i = \frac{\alpha_i C_i}{\beta}. \tag{6}$$

From (4) and (5) we see that the diagonal terms of the matrix of (5) are all 1, whereas the members not on the diagonal are smaller than 1.

* Received Nov. 1, 1946. Communication No. 1123 from the Kodak Research Laboratories.

The senior author now suggests an ordering of the vectors in such a way that the principal subdeterminants* have a maximum value (which will always be less than 1, since k independent edges of the parallelopiped formed by the unit vectors $\bar{s}_i$ lie inside the unit sphere). If the determinant of the problem is small, this means that the volume of the parallelopiped is small or that the vectors $\bar{a}_i$ are not "too independent," i.e., that one or more of the vectors can be approximated by a linear combination of the others.

If we denote by a_{ij} the quantity, $s_i s_j$, by $c_i^{(0)}$ the quantity ts_i, we may write (5) in the form

$$\begin{aligned} a_{11}\gamma_1 + a_{12}\gamma_2 + \cdots + a_{1k}\gamma_k &= c_1^{(0)}, \\ \cdots\cdots\cdots\cdots\cdots\cdots & \\ a_{k1}\gamma_1 + a_{k2}\gamma_2 + \cdots + a_{kk}\gamma_k &= c_k^{(0)}. \end{aligned} \tag{7}$$

The method used will be this: We choose an arbitrary number, say r, where $r<k$, and find the r-rowed principal subdeterminant which has the largest value. We set the $(k-r)$ variables not involved equal to zero, and solve the r equations for the r unknown variables, obtaining the values $\gamma_{j,1}=\bar{\gamma}_{1,1}\cdots\bar{\gamma}_{r,1},\gamma_{j,1}=0$, $j>r$. Substituting these for the variables $\gamma_{1,1}\cdots\gamma_{r,1}$ in the $k-r$ remaining equations, we have

$$\begin{aligned} a_{1,r+1}\gamma_{r+1} + \cdots + a_{1,k}\gamma_k &= c_1^{(1)}, \\ a_{k,r+1}\gamma_{r+1} + \cdots + a_{k,k}\gamma_k &= c_k^{(1)} \end{aligned} \tag{8}$$

where $c_i^{(1)}=c_i^{(0)}-\sum_{j=1}^{r}a_{i,j}\bar{\gamma}_{j,i}$. At this point we have two possibilities. Either the constants $c_i^{(1)}$ are very near to zero, in which case we have that the solution of (8) is given by $\gamma_j=0$, $j>r$, so that a reasonably accurate solution of (7) and thus of (5) is given by

$$\gamma_{i,1} = \bar{\gamma}_{i,1} \quad (i = 1, 2, \cdots, r); \qquad \gamma_{i,1} = 0 \quad (i = r+1, \cdots, k). \tag{9}$$

In the event that they are not sufficiently small, we take the largest $(r+1)$-rowed principal subdeterminant of (7), equate the remaining $(k-r-1)$ variables not involved to zero, and proceed as before, obtaining $(k-r-1)$ equations in $(k-r-1)$ unknowns:

$$\sum_{\lambda=1}^{k-r-1} \mu_{i\lambda}\gamma_{j\lambda} = c_i^{(2)}, \qquad (i = 1, 2, \cdots, k-r-1), \tag{10}$$

where the meaning of the terms $\mu_{i\lambda}$ and $c_i^{(2)}$ is clear. Should the constants $c_i^{(2)}$ be sufficiently small, the approximate solution is given by

$$\gamma_i = \bar{\gamma}_i \quad (i = 1, 2, \cdots, r+1); \qquad \gamma_i = 0, \quad (i = r+2, \cdots, k). \tag{11}$$

One must keep in mind that as larger subdeterminants are taken, their values get smaller, and that, of course, as r gets nearer k the $c_i^{(n)}$ become smaller, and vanish when $r=k$. There will usually be a place somewhere in this process, however, where sufficiently good accuracy will be achieved. If greater accuracy is desired, one must

* Subdeterminants whose principal diagonals form part of the principal diagonal of the determinant of the system (5).

attempt one of the ordinary approximation methods, in particular, an iteration procedure.

We present as an example a set of six equations with six unknowns. Since the calculation of the maximum subdeterminant is very laborious, we have replaced it by the less laborious, but theoretically less satisfying, method of ordering the vector so that the sum of the squares of the elements outside the principal subdeterminant is as small as possible.

The system is represented by the following matrix.

I

6.35083641	3.38272218	1.83936868	1.02023449	0.57681127	0.33173876	2.27198287
	1.83271729	1.01325021	0.57097916	0.32757965	0.19093008	1.26347780
		0.56915637	0.32551164	0.18928551	0.11166820	0.71717786
			0.18871015	0.11107961	0.06624255	0.41473305
				0.06609478	0.03979216	0.24391707
					0.02415650	0.14552476

This matrix is reduced to unit diagonal and reordered as described above (in this example, the ordering remains the same). The elements of the resulting matrix are the light-face elements of the following matrix.

II

1.00000000	0.99152377	0.96747004	0.93193774	0.89029657	0.84696167	0.90154952	**0.90154952**
	1.00000000	0.99209478	0.97090085	0.94120796	0.90742269	0.93329702	
−0.99152377	0.01688061	**0.03282524**	**0.04686243**	**0.05845775**	**0.06764006**	**0.03938924**	**2.33340146**
		1.00000000	0.99323746	0.97592846	0.95235029	0.95062903	
−0.96747004	−1.94455295	**0.00017130**	**0.00048914**	**0.00091901**	**0.00141057**	**0.00181242**	**10.58038529**
			1.00000000	0.99461070	0.98112063	0.95470899	
−0.93193774	−2.77610999	−2.85545826	0.00000007				
				1.00000000	0.99585694	0.94876491	
−0.89029657	−3.46301170	−5.36491535					
					1.00000000	0.93631051	
−0.84696167	−4.00696776	−8.23450088					

The bold-face elements are found as follows: The first bold-face row contains the determinants of the second order found from the first and second rows. Any other bold-face elements above the stairway dividing line are formed by adding to the element immediately above it the product of the rest of the column above it and the row to its left, e.g., $0.99323748-0.93193775\times0.96747003-0.04686244\times1.94455183=0.00048919$. The first column to the left of the stairway dividing is the negative of first row, the second column is the negative of the first bold-face row divided by its first number, and so on. The process is continued until one of the diagonal elements becomes very small. The solution of the equations, omitting the last and successive variables, is found in the usual manner from the bold-face column at the extreme right. This column is the ratio of the last number of each row to the first number of each row.

We find in our case

$$\gamma_1 = 8.75144919, \qquad \gamma_2 = -18.24071797, \qquad \gamma_3 = 10.58038529.$$

The three remaining unknowns are found by distributing the residuals among all six equations by least squares. When this is done, we find for our original unknowns:

$$C_1 = 3.4727, \qquad C_2 = -13.4739, \qquad C_3 = 14.0244,$$
$$C_4 = -0.0003, \qquad C_5 = 0.0002, \qquad C_6 = 0.0001.$$

This method has been applied successfully to least-squares solutions in geometrical optics and to the colorimetric problem of finding polynomial reflection curves which will yield a prescribed set of tristimulus values under a given illuminant and fit other prescribed conditions.

In the latter problem, the small unknowns must be neglected prior to symmetrization and reduction to unit diagonal.

NEW FORMULATIONS OF THE EQUATIONS FOR COMPRESSIBLE FLOW*

By B. L. HICKS (*Ballistic Research Laboratories, Aberdeen Proving Ground*)
P. E. GUENTHER (*Case School of Applied Science*) AND R. H. WASSERMAN (*University of Chicago*)

Introduction. A prominent aerodynamic effect of combustion in a moving gas stream is an alteration of the flow pattern owing to heat release within the fluid. This alteration occurs not only in the immediate neighborhood of heat sources but also downstream where the entropy and stagnation temperature vary from one streamline to another. As a background for combustion research, appropriate descriptions of these altered flow patterns have been investigated. This paper considers the downstream patterns, which are restricted to be the adiabatic and steady flows of an inviscid fluid. In a second paper,[1] diabatic (i.e., non-adiabatic) flows will be discussed.

Since one-dimensional flow theory[2,3] can be considerably condensed by use of the local Mach number M, it was natural to seek a corresponding condensation with the help of the Mach vector

$$\mathbf{M} = M\mathbf{V}/V$$

and the Crocco vector

$$\mathbf{W} = \mathbf{V}/V_t$$

in which V_t is the variable limiting velocity at each point of the fluid. The introduction of Mach and Crocco vectors into the compressible flow equations sufficiently simplified or altered their form that a number of further investigations were suggested including those of diabatic flow.

* Received Feb. 12, 1947. This paper is a revised report of theoretical work performed by the authors at the Cleveland Laboratory of the National Advisory Committee for Aeronautics in 1943–45.

[1] B. L. Hicks, *Diabatic flow of a compressible fluid*, submitted to Quarterly of Applied Mathematics.

[2] Neil P. Bailey, *The thermodynamics of air at high velocities*, Journ. Aero. Sci. **11**, 227–238 (1944).

[3] B. L. Hicks, D. J. Montgomery, and R. H. Wasserman, *The one dimensional theory of steady compressible fluid flow in ducts with friction and heat addition*, NACA TN, 1947.

1. Development of the new equations. The basis for our development is the following set of equations:

$$c_v\rho\mathbf{V}\cdot\nabla T + p\nabla\cdot\mathbf{V} = 0, \tag{1.1}$$

$$\rho\mathbf{V}\cdot\nabla\mathbf{V} + \nabla p = 0, \tag{1.2}$$

$$\nabla\cdot\rho\mathbf{V} = 0, \tag{1.3}$$

$$p = R\rho T. \tag{1.4}$$

It is appropriate to include (1.1), an expression of energy conservation, because the entropy S and stagnation temperature T_t are not constant throughout the field of flow. The Bernoulli relation,

$$c_pT_t = c_pT + \tfrac{1}{2}V^2, \tag{1.5}$$

then defines the stagnation temperature. In terms of M and W, equation (1.5) becomes

$$(T_t/T) = 1 + (\gamma - 1)M^2/2 = (1 - W^2)^{-1}. \tag{1.6}$$

The Mach vector equations are derived from (1.1), (1.2), (1.3), (1.4) by the substitution $\mathbf{V} = \sqrt{\gamma RT}\,\mathbf{M}$ and elimination of ρ and T. The new equation of motion resulting is

$$\mathbf{M}\cdot\nabla\mathbf{M} - \frac{\gamma - 1}{\gamma + 1}\mathbf{M}\nabla\cdot\mathbf{M} + \frac{1}{\gamma}\nabla\log p = 0, \tag{1.7}$$

and the new equation of continuity

$$\nabla\cdot\left(1 + \frac{\gamma - 1}{2}M^2\right)^{-(\gamma+1)/2(\gamma-1)}\mathbf{M} = 0. \tag{1.8}$$

The equations of motion and continuity in terms of the Crocco vector are derived similarly, the initial transformations being $\mathbf{V} = V_t\mathbf{W}$ and $T = V_t^2(1 - W^2)/2c_p$. It is in addition necessary to use the fact that T_t and therefore also V_t are constant on streamlines for adiabatic, inviscid flow (cf., for example, Vaszonyi's proof[4]). The Crocco vector equations are then

$$\mathbf{W}\cdot\nabla\mathbf{W} + \frac{\gamma - 1}{2\gamma}(1 - W^2)\nabla\log p = 0, \tag{1.9}$$

$$\nabla\cdot(1 - W^2)^{1/\gamma-1}\mathbf{W} = 0. \tag{1.10}$$

2. General comments. The local behavior of a compressible fluid is characterized by its Mach number. Therefore the vectors $\mathbf{M}$ and $\mathbf{W}$, from whose magnitudes M can be determined, supply more useful descriptions of a flow than the vector $\mathbf{V}$ from whose magnitude M cannot be computed without knowledge of V_t. The equations possess equal generality in all three languages but are simpler in $\mathbf{M}$ and $\mathbf{W}$ language for these representations have permitted elimination of all but one vector variable from the equation of continuity and elimination of all but this vector and p from the equation of motion. Such reduction in the number of independent variables has heretofore been accomplished only for iso-energetic flow.

[4] A. Vazsonyi, *On rotational gas flows*, Quarterly Appl. Math. **3**, 29–37 (1945).

Rotational flows are most aptly described by the **W** language; the added term $\mathbf{M}\nabla\cdot\mathbf{M}$ in the **M** equation of motion brings with it no advantage. (The **W** language is still the best when *diabatic* rotational flows are considered.) However when either **M** or **W** is irrotational, the physical characteristics of the resulting flows make it appropriate to retain each language.

3. Geometry of stream tubes and transition through the speed of sound. In the one-dimensional approximation, isentropic compressible flows have minimum flow area at sonic velocity. The same property has been suggested by Prandtl[5] for irrotational three-dimensional flows. We shall prove that this property is valid for all continuous, steady adiabatic flows when an appropriate description of stream-tube area is introduced.

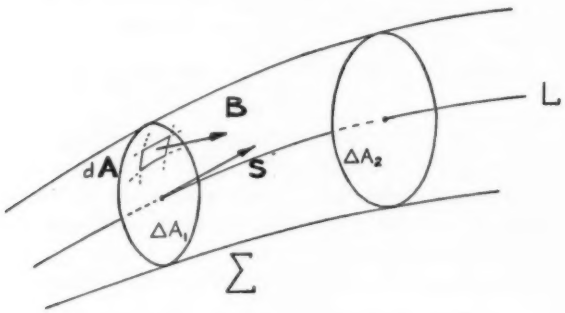

FIG. 1.

Let **B** be any continuous and solenoidal vector parallel to **M**. Consider any two plane sections ΔA_1, ΔA_2 of the stream-tube Σ which are normal to the streamline L contained in Σ. Since $\int_{\Delta A_1}\mathbf{B}\cdot d\mathbf{A}_1=\int_{\Delta A_2}\mathbf{B}\cdot d\mathbf{A}_2=\int_{\Delta A}B_n dA$ a position (m) can be found for each section such that $[B_n(m)\Delta A]_1=[B_n(m)\Delta A]_2$ or

$$\frac{\partial}{\partial s}\log B_n(m)+\frac{1}{\Delta A}\frac{\partial \Delta A}{\partial s}=0, \tag{3.1}$$

in which $\partial/\partial s$ denotes differentiation along L. As $\Delta A\to 0$, $B_n(m)\to B$, the magnitude of B on L. Also if $\mathbf{s}=\mathbf{B}/B$ is the unit vector in the direction of flow

$$\nabla\cdot\mathbf{s}=\lim_{\Delta A\to 0}\frac{1}{\Delta A}\frac{\partial \Delta A}{\partial s}. \tag{3.2}$$

The fractional rate of change of stream-tube area with respect to streamline arc is therefore measured by $\nabla\cdot\mathbf{s}$ for flow in both two and three dimensions. Similar relations derived geometrically have been used by von Kármán[6] for two-dimensional flows.

The continuity equation in **M** language can be rewritten in terms of $\nabla\cdot\mathbf{s}$ as

$$M^{-1}\left(1+\frac{\gamma-1}{2}M^2\right)^{-1}(1-M^2)\frac{\partial M}{\partial s}+\nabla\cdot\mathbf{s}=0. \tag{3.3}$$

For $\nabla\cdot\mathbf{s}>0$, $|1-M|$ cannot decrease but for $\nabla\cdot\mathbf{s}<0$, $|1-M|$ must decrease in the di-

[5] L. Prandtl, *General considerations on the flow of compressible fluids*, NACA TM No. 805, 1936.

[6] Th. von Kármán, *Compressibility effects in aerodynamics*, Jour. Aero. Sci. **8**, 337–356 (1941).

rection of flow i.e. sonic velocity is approached. As $M \to 1$ however, $\partial M/\partial s$ diverges like $-(\gamma+1)(\nabla\cdot\mathbf{s})/4(1-M)$ and $\partial p/\partial s$ diverges similarly. Consequently if M is to pass through the value unity continuously, $\nabla\cdot\mathbf{s}$ must be zero there. Consideration of the signs shows that the indicated extremum corresponds to a minimum of stream-tube area at $M=1$.

4. Thermodynamic considerations. The equation

$$\mathbf{V} \times (\nabla \times \mathbf{V}) = c_p \nabla T_t - T\nabla S \tag{4.1}$$

which is similar to Vazsonyi's,[4] contains the thermodynamic implications of the **V** equations. (Less general equations than (4.1) have also been derived by Crocco,[7] Tollmien[8] and Emmons.[9]) The corresponding equation in terms of **W** and the stagnation pressure p_t is easily obtained from (1.9)

$$\mathbf{W} \times (\nabla \times \mathbf{W}) = \frac{\gamma - 1}{2\gamma}(1 - W^2)\nabla \log p_t \tag{4.2}$$

showing that there is one fundamental thermodynamic quantity in adiabatic flow, namely p_t, and not two separate ones S and T_t. Thus for a given $p_t = p_t'$ distribution, the same **W** (or **M** but not **V**) field obtains no matter what the variation of S and T_t so long as they combine in the form $[T_t \exp(-S/c_p)]^{\gamma/\gamma-1}$ to give p_t'. There seems to be no corresponding theorem for equation (4.1).

Differentiation of (4.1) yields a differential form of Bjerknes' theorem[8]

$$\nabla \times [\mathbf{V} \times (\nabla \times \mathbf{V})] = -\nabla T \times \nabla S. \tag{4.3}$$

Differentiation of (4.2) results in Crocco's equation[7] containing **W** only:

$$\nabla \times \left[\frac{\mathbf{W} \times (\nabla \times \mathbf{W})}{1 - W^2}\right] = 0. \tag{4.4}$$

Equations (1.10), (4.3) and (4.4) were originally derived for isoenergetic flow, a restriction now seen to be unnecessary.

5. Irrotational flows. We now consider in turn the consequences of irrotationality of **M**, **W** and **V** fields. If $\mathbf{M} = \nabla\varphi_M$, an integrability condition

$$\nabla^2\varphi_M = f(\varphi_M) \tag{5.1}$$

must be satisfied if the **M** equation of motion (1.7) is to be integrable. (A similar equation for the stream function occurs in theory of rotational incompressible flow.)[10]. Since the potential function must also satisfy the continuity equation (1.8) it is to be expected that the manifold of permissible functions φ_M will be somewhat restricted. This is illustrated by the fact that $p \exp(\gamma M^2/2)$ is a function of φ_M alone, or that M, φ_M and p_t are related by the expression

[7] L. Crocco, *Eine neue Stromfunktion für die Erforschung der Bewegung der Gase mit Rotation*, ZaMM 17, 1–7 (1937).

[8] W. Tollmien, *Ein Wirbelsatz für stationäre isoenergetische Gasströmungen*, Luftfahrtforschung 19, 145–147 (1942), British R.T.P. Trans. No. 1744, Ministry Aircraft Prod.

[9] H. W. Emmons, *The numerical solution of compressible fluid flow problems*, NACA TN No. 932, 1944.

[10] M. Lagally, *Ideale Flüssigkeiten*, Handbuch der Physik, Julius Springer (Berlin), vol. 7, ch. I, Art. 19, p. 29, and Art. 32, p. 49.

$$\Xi(M) = \Phi(\varphi_M) - \frac{1}{\gamma} \log p_t, \tag{5.2}$$

in which

$$\Xi(M) = -\frac{\gamma-1}{2} \int M^3 \left(1 + \frac{\gamma-1}{2} M^2\right)^{-1} dM,$$

$$\Phi(\varphi_M) = \frac{\gamma-1}{\gamma+1} \int f(\varphi_M) d\varphi_M. \tag{5.3}$$

(In uniplanar flow, explicit formulae for potential and stream functions can be derived for both irrotational **M** and some irrotational **W** flows. We have not ascertained whether these formulae represent more than the well-known elementary radial and vortex flows.)

If $\mathbf{W} = \nabla\varphi_W$, then $\nabla p_t = 0$, a case discussed by Vazsonyi[4] in the less appropriate **V** language. Flows with constant stagnation pressure but variable entropy and stagnation temperature from one streamline to another may be of immediate practical interest. For example the effect of variable chamber gas temperature upon thrust coefficient of a rocket with "perfect" nozzle can be computed. This calculation is possible because irrotational **V** flow is the special case of irrotational **W** flow which occurs when not only p_t but also S and therefore T_t are constant throughout the flow ($\nabla S = \nabla T_t = 0$ is implied by $\mathbf{V} = \nabla\varphi_V$). Accordingly the same partial differential equation (derived from the continuity equation (1.10)) is satisfied by both φ_W and its isentropic form φ_V,

$$\sum_{i=1}^{3} \left[1 - W^2 - \frac{2}{\gamma-1}\left(\frac{\partial\varphi_w}{\partial x_i}\right)^2\right] \frac{\partial^2\varphi_w}{\partial x_i^2} - \frac{4}{\gamma-1} \sum_{i>j=1}^{3} \frac{\partial\varphi_w}{\partial x_i} \frac{\partial\varphi_w}{\partial x_j} \frac{\partial^2\varphi_w}{\partial x_i \partial x_j} = 0. \tag{5.4}$$

LINEARIZATION OF SOLUTIONS IN SUPERSONIC FLOW*

By JOHN W. TUKEY (*Bell Telephone Laboratories and Princeton University*)

1. Introduction. The equations governing flow at supersonic speeds are believed to be well known, but the difficulties of calculating exact solutions are so great that approximate solutions are the aim of the present and foreseeable future. Two main approaches to such approximate solutions are commonly considered:

(a) the calculation of numerical approximations by high-speed calculators,

(b) simplification of the equations and explicit solution of the simplified equations.

Of course the first route requires the high-speed calculating machines which are now in sight, but not yet available.

It is the purpose of this note to propose and exemplify a third approach which may partly replace (b) and frequently supplement (a), namely

(c) simplification of the dependence of the answers on one or more parameters.

2. Is linearization of solutions mathematically justified? The first objection which many would raise against "linearizing the solutions" as a substitute for "linearizing

* Received Feb. 15, 1947.

the equations" is its mathematical justification. The justification of the former would seem to be at least as good as the latter. In order to make this clear, let us look at the situation in more detail.

A problem is believed to be adequately stated by a complex set of equations, which we may symbolize by

$$F(x, \lambda) = 0, \tag{1}$$

where λ is a parameter. The classical procedure is to replace the complex set of equations by a simpler set, frequently composed of linear equations, which we may write

$$L(x, \lambda) = 0, \tag{2}$$

and then to solve *these* equations *exactly* to obtain

$$x = g(\lambda).$$

We must interpret this solution as an approximate solution of the original equations. What justification might we have? The only possibilities seem to be

(a) a theorem that the solutions of (2) are close to the solutions of (1),
(b) empirical evidence that, in particular cases, the solutions of (2) are close to the solutions of (1),
(c) mathematical intuition,
(d) physical intuition.

The author knows of no case in supersonic flow where the very desirable justification of type (a) exists.

On the other hand, suppose that (1) can be solved for, say, two values of λ, so that we know x_1 and x_2, where

$$F(x_1, \lambda_1) = 0 = F(x_2, \lambda_2), \tag{3}$$

and suppose further that we believe that we can approximate the dependence of x on λ over a certain range by one of, say, a two-parameter family of functions,

$$x(\lambda) = h(\lambda; A, B). \tag{4}$$

We can then use the two known cases to determine the parameters and find an approximate solution

$$x = h(\lambda; A_0, B_0) \tag{5}$$

of (1). What justification might we have? The only possibilities seem to be

(a) a theorem that the solution of (1) depends on λ in a way close to one of the functions (4),
(b) empirical evidence that, in particular cases, the solutions of (1) do behave like one of the functions (4),
(c) mathematical intuition,
(d) physical intuition.

The author knows of no case in supersonic flow where the very desirable justification of type (a) exists.

Thus the two ways of "linearizing" seem equally justified. The first appeals to some because the solution used is *exactly* a solution of (2). There is no reason, how-

ever, why this should be more esthetic than the fact that the other approximation is in exactly the form 4.

3. A linearization for the region of attacked shock waves. In the range of supersonic speeds safely away from sonic or hypersonic speeds, there seems to be some evidence that a linearization of the form

$$Ap + Bq, \tag{5}$$

where q, the dynamic pressure, is given by

$$q = \frac{\gamma}{2} M^2 p, \tag{6}$$

applies approximately to many *averaged quantities*, such as lifts, drags, etc., for well streamlined bodies. Some empirical evidence is presented in the next section. It is believed that, when the two points needed to determine A and B are known from theory or experiment, the use of this linearization will frequently give results preferable to those obtained by solving a linearized problem.

4. The flow around an axially-symmetrical cone. The theoretical problem of flow around a cone was solved by Taylor and Maccoll[1] who computed the drag or pressure coefficient for several cases. These calculations were refined and extended by Maccoll.[2]

A linearized theory was previously presented by v. Kármán and Moore.[3]

The present paragraph is concerned with the empirical linearization

$$C_p \doteq \left(0.083 + \frac{0.096}{M^2}\right)\left(\frac{\theta_c}{10^\circ}\right)^{1.69}, \qquad (\theta_c = \text{half angle of cone}). \tag{*}$$

The form and constants of this approximation are purely empirical, and are principally based on Maccoll's theoretical values.

Figure 4 on p. 309 of the paper by Taylor and Maccoll[1] shows that the linearized equations give rise to errors of 10 to 25 per cent over substantial portions of the range, while the following table shows that the empirical formula gives agreement to better than 5 per cent over a range of half angles from 10° to 55° (Maccoll estimates 57.4° as the maximum half angle for conical flow) and over a range of Mach numbers for each angle extending down to within about 0.1 of the minimum Mach number for conical flow (attached shock-wave).

The fit is not perfect, and it is clear that the empirical formula cannot replace the exact solution for all purposes. However, as a means of interpolating between known values, it clearly serves a useful purpose.

Summary. There is theoretical and empirical justification for "linearizing the answer" rather than "linearizing the equations" in some fields of supersonic flow.

---

[1] G. I. Taylor and J. W. Maccoll, *The air pressure on a cone moving at high speeds*, Proc. Roy. Soc. London (A) **139**, 278–311 (1933).

[2] J. W. Maccoll, *The conical shock wave formed by a cone moving at a high speed*, Proc. Roy. Soc. London (A) **159**, 459–472 (1937).

[3] Th. von Kármán and N. B. Moore, *Resistance of slender bodies moving with supersonic velocities, with special reference to projectiles*, Tr. Amer. Soc. Mech. Eng. **54**, 303–310 (1932); for further discussion see Taylor and Maccoll, *loc. cit.*, p. 30.

TABLE 1. Supersonic Pressure Coefficient of Unyawed cone

Semiangle of Cone (Min. Mach No.[2])	Mach Number	Pressure Coefficient[1] Quadrature[3]	Pressure Coefficient[1] Empirical[4]	Per Cent Deviation
10°	1.091	.156[5]	.161	+5
(1.04)	1.403	.130	.131	+1
	1.81	.108	.112	+4
	2.39	.096	.100	+4
	3.33	.090	.092	+2
	5.46	.076	.086	+13
20°	1.314	.446	.447	+0
(1.18)	1.555	.374	.394	+5
	1.65	.356	.382	+7
	2.13	.318	.333	+5
	2.86	.296	.306	+3
	4.17	.268	.286	+7
	9.74	.256	.271	+6
30°	1.514	.814	.800	−2
(1.46)	1.639	.730	.760	+4
	2.05	.636	.678	+7
	2.71	.598	.614	+3
	3.16	.570	.593	+4
	3.85	.556	.572	+3
	7.84	.524	.542	+3
40°	2.03	1.098	1.106	+1
(1.95)	2.22	1.036	1.067	+3
	2.84	0.948	0.988	+4
	4.20	0.904	0.921	+0
	5.86	0.890	0.893	
45°	2.38	1.316	1.270	−4
(2.3)	2.72	1.182	1.219	+4
	3.76	1.104	1.140	+4
	4.89	1.084	1.105	+2
	8.01	1.068	1.073	+1
50°	3.15	1.440	1.407	−2
(3.1)	3.33	1.374	1.391	+1
	3.81	1.330	1.360	+2
	4.75	1.296	1.325	+2
	7.11	1.266	1.289	+2
55°	5.59	1.552	1.535	−1
(5.5)	6.38	1.516	1.522	0
	7.56	1.496	1.510	+1

[1] Given in American units—Maccoll's values have been doubled.
[2] Minimum for conical flow and attached shock wave.
[3] Taken from Maccoll, Proc. Roy. Soc. **159** (1937) pp. 461 and **469**, also **139** (1933) p. 285.
[4] Calculated by formula (*) above.
[5] This value is undoubtedly too low by 10 to 20 per cent.

The linearization

$$(\text{drag, lift, etc.}) = Ap + Bq$$

is proposed for use in the region of attached shock waves.

BOOK REVIEWS

Tables of spherical Bessel functions. Prepared by the Mathematical Tables Project, National Bureau of Standards. Vol. 1. Columbia University Press, New York, 1947. xxviii+375 pp. $7.50.

In this volume the functions $\sqrt{\pi/2x}J_\nu(x)$ are tabulated for the half-odd-integer values of ν, $-27/2 \leqq \nu \leqq 27/2$. For $0 \leqq x \leqq 10.0$, the functions are tabulated to eight or more significant figures at intervals of .010. For $10 \leqq x \leqq 25$, the tables contain seven significant figures and the argument interval is .10. A table of the function $x^{-\nu+1/2}\sqrt{\pi/2x}J_\nu(x)$ is also included for argument intervals of .01. The range of argument is: for $\nu = -\frac{1}{2}$, $\pm 5/2$, $\pm 7/2$: $0 \leqq x \leqq 1$; for $\nu = \pm 3/2$: $0 \leqq x \leqq .50$; for $\nu = \pm 9/2$, $\pm 11/2$, $\cdots$, $\pm 15/2$: $0 \leqq x \leqq 1.5$; for $\nu = \pm 17/2$, $\cdots$, $23/2$: $0 \leqq x \leqq 2$; and for $\nu = \pm 25/2$, $\pm 27/2$: $0 \leqq x \leqq 2.5$.

G. F. CARRIER

Introduction to aerodynamics of a compressible fluid. By Hans Wolfgang Liepmann and Allen E. Puckett. Galcit Aeronautical Series. John Wiley & Sons, Inc., New York, 1947. ix+262 pp. $4.00.

The better known works on the aerodynamics of compressible fluids are 1) Ackeret's and 2) Busemann's articles in the Handbuch der Physik and the Handbuch der Experimentalphysik; 3) Bateman's article in the Report of the Committee on Hydrodynamics of the National Research Council; 4) Taylor and Maccoll's article in Durand's Aerodynamic Theory; 5) papers collected in the Proceedings of the Fifth Volta Congress and 6) Sauer's rather recent book. The first three, although excellent summaries of information up to the dates of their publication are now outdated. The fourth is masterly in its choice of subjects and exposition but may be too difficult for a beginner in the subject due to its briefness. The fifth contains truly a wealth of good papers covering the whole field, but unfortunately it is out of print. The sixth and last one is written by a mathematician from the point of view of a mathematician.

Therefore for aeronautical engineers now entering upon a study of aerodynamics of compressible fluids, there was a great need for an introductory book with a clear exposition of the physical aspects and the recent advancement of the subject. This need is believed to be admirably satisfied by Liepmann and Puckett's book. Of course, as the title indicates, the book is not a compendium of all available data on compressible flows, and for engineers who are desperately looking for design information, this book will be a disappointment. However, for those who are interested in a clear understanding of the phenomena of compressible flow, the outstanding feature of this book is the constant emphasis on the interplay between theory and physical facts. The authors always try to present a description before going into a calculation. Thus the gist of a problem is not lost in a maze of equations. It also is evident that the authors are always aware of the limitations imposed by simplifying assumptions which are, of course, necessary in order to make the problem tractable, and such limitations are always pointed out in due time. The reader is thus saved from the various pitfalls in applying the results of a theory. The authors' research efforts in compressible flow are also evident in the book because not only are known facts expounded, but difficult questions in such subjects of current interest as condensation shock, Mach reflection, influence of viscosity in transonic flows are also discussed.

However there are occasions in the book where the authors leave an inquisitive mind unsatisfied. For instance, the treatment in Part II on two- and three-dimensional motions is uneven in its pace.—While great details of computations are given in deriving the basic perturbation equations, no method of solving these equations or example is given, so that reader could have a more concrete understanding of the principles involved. Then for condensation shocks, there is the question of the relative importance of dust

nucleii and nucleii formed by statistical fluctuations from thermodynamical equilibrium, and the "theoretical" possibility of the unobserved condensation shock to subsonic velocities. The authors' discussion on these questions are vague and indecisive. The form of expanison of potential into a series of thickness ratio of the body in Chapter 8 is, of course, incorrect for any three-dimensional axially symmetric flow and for any two-dimensional flow with a stagnation point. However the reviewer does not consider these statements as criticisms of this excellent book.—If the book were without such defects, then it would be a book by a master. But such definitive treatise on any one subject can be expected to appear, perhaps, only once in fifty years.

H. S. TSIEN

The theory of mathematical machines. By Francis J. Murray. King's Crown Press, New York, 1947. vii+116 pp. $3.00.

This book discusses the operation of many basic units used in mathematical machines. The interest is in the basic devices themselves rather than in the problems arising when they are interconnected.

The four chapters of the book are entitled Digital Machines, Continuous Operators, The Solution of Equations, and Mathematical Instruments. Each of these is a series of discussions of various devices. A typical discussion contains a description of the physical appearance together with a brief analysis of the operation. Both mechanical and electrical schemes are discussed under the assumption that the reader has a reasonable acquaintance with the operation of simple machine components and with elementary electrical, and particularly electronic, principles.

The subject matter is a collection of material which for the most part is widely scattered in the literature of various fields. For this reason, the book would be extremely interesting to anyone wishing to obtain an acquaintance with the subject. The treatment of analogue devices using electrical networks and mechanical devices is fairly complete and quite interesting—particularly that part dealing with the author's own work. On the other hand, more space might have been devoted to digital techniques and less to the discussion of engineering design features which are occasionally discussed at some considerable length.

The book is issued by the King's Crown Press in loose leaf form "—as submitted by the author, without the usual editorial attention of the Columbia University Press," in an effort to minimize the cost to the buyer. It is this reviewer's opinion that further editorial work would have been well warranted since there are many minor errors and vague statements, and the lack of numbered references to diagrams makes it difficult at times to follow the technical arguments.

J. A. KRUMHANSL

Methods of mathematical physics. By Harold Jeffries and Bertha S. Jeffries. Cambridge: At the University Press, 1946. vii+679 pp. $15.00.

"Methods of Mathematical Physics" is a very comprehensive work dealing with the mathematical techniques of classical and modern physics. Despite the fact that the book is primarily intended for those who apply mathematics, the presentation is sufficiently careful to satisfy reasonable requirements as to rigor.

The introductory chapter provides a brief but satisfactory background in the theory of functions of a real variable and the book continues through a detailed discussion of such topics as matrices, tensors, function theory and related subjects, asymptotic expansions, the classical equations (equations of heat conduction, wave equation, Laplace equation, etc.), and special functions.

As in any book of this scope, there are a few notable omissions. The Fourier analysis might have been presented in more general form; the isoperimetric problems and direct methods of the variational calculus are absent, and, in the chapter on numerical methods, a wealth of detail on interpolation formuli leaves no room for many valuable topics, e.g. the Rayleigh-Ritz procedure and the techniques for finding eigenvalues of matrices. It also seems to the reviewer that a more fortunate presentation of the Laplace Transform techniques might have been used. However, these latter remarks are greatly outweighed by the over-all value of the book which should find a place as an excellent reference for the mathematical physicist.

G. F. CARRIER

Matrix and tensor calculus, with applications to mechanics, elasticity and aeronautics. By Aristotle D. Michal. Galcit Aeronautical Series. John Wiley & Sons, Inc. New York, 1947. xiii+132 pp. $3.00.

The purpose of this book is to furnish engineering students (especially professional engineers) with an introduction to the theory and applications of the matrix and tensor analysis. In view of this purpose and the vast literature in each of these subjects, the author develops only: (1) those aspects of the theory of matrices which are of immediate application to the problems of vibration theory; (2) those elements of tensor analysis which suffice for a tensor formulation of the Lagrange, elasticity and hydrodynamical equations. The mathematical theory is presented in concise form and some applications of the theory to practice are clearly indicated. The problems supplement the material of the text and, though few in number, should provide the engineer with some of the necessary practice in mastering the new tools.

Some indication of the material contained in the book may be obtained by a brief summary of the topics covered. Part I deals with such topics as addition, multiplication, differentiation and integration of matrices, the Cayley-Hamilton theorem as applied to the calculation of the inverse of a given matrix, systems of matrix differential equations and the solution of oscillation problems for characteristic frequencies and amplitudes. The problems contain some applications to flutter theory. Most of Part II is concerned with tensor analysis in Euclidean three space. Scalers, vectors, tensors (contravariant and covariant) are defined. As examples of tensors, the author studies the metric and strain tensors. The Laplace, Poisson, and wave equations are derived in general curvilinear coordinates. The related problem of the connection between the modern and classical theories of tensors in orthogonal curvilinear coordinates (see L. Brillouin, Les Tenseurs, Dover, New York, 1946, p. 114) is not discussed in the text. However, the author furnishes a detailed and interesting tensor treatment of elasticity theory. The last two chapters of the book deal with the tensor analysis of n-dimensional Riemann space. Geodesics and the Lagrange equations of motion are discussed. Finally, the Navier-Stokes equations for viscous fluids are formulated in tensor form and a treatment of the boundary layer problem by a perturbation method is indicated (the author gives credit to Dr. C. C. Lin for this latter work).

In view of the many uses of matrix theory and the value of tensor analysis in gaining an insight into the geometric structure of partial differential equations, it has become essential that the engineer obtain some grasp of these subjects. This text should furnish the engineer with a good introduction to these fields.

N. COBURN

Introduction to the theory of equations. Second edition. By Lois W. Griffiths, John Wiley and Sons, Inc., New York; Chapman and Hall, Ltd., London, 1947. 278 pp. $3.50

The usual topics are treated here with unusual attention to detail. Each general theorem is preceded by illustrative special cases, and numerous problems are included. Proofs of the main theorems are accomplished by gradually building up the desired statements out of simpler ones. The author is as fully concerned with teaching the student how to prove these theorems as with the theorems themselves. The result is a somewhat long but very precise text which might best serve the student who intends to continue in mathematics.

H. J. GREENBERG

Statistical analysis in biology. Second Edition. By K. Mather. Interscience Publishers, Inc., New York, 1947. 267 pp. $5.00.

This book is written for biologists with a rather weak background in mathematics. This aim has been so well carried out that many mathematicians will find the lack of analytical short-cuts annoying. They have no ground for complaint. The book can be recommended to biologists as one of a small number with modern techniques and relatively little mathematics. It can be recommended to mathematicians interested in how statistics is actually used—in biology at large, and more especially, in the author's field of genetics.

The second edition differs from the first by the correction of some errors and the addition of a thir-

teenth chapter on the angular and probit transformations. The book's special features include (*i*) the only available careful, elementary account of single degrees of freedom and (*ii*) the use of small sample tests before their large sample brethren. The book seems remarkably free of typographical errors, the reviewer noting only "ten" for "nine" near the top of pages 228 and 231.

Although it is a revised edition, there are ten or a dozen points with which the reviewer differs. Some are errors of fact, others are differences in philosophy. Some of the most important seem to have arisen by incautious reading of R. A. Fisher—paying heed to what Fisher's words seem to mean rather than to what Fisher was very careful to say.

The "rigour" of statistical inferences is emphasized at page 11, without any warning that the exactness of these Fisherian procedures (which the reviewer uses himself) depends on the exact normality of the distributions found in practice—something difficult if not impossible to verify by experiment.

A minor point on page 35 concerns kurtosis. Here overprevalence in the center is given equal weight with overprevalence in the tails as a cause of kurtosis. This is only half the classical error, which emphasized the center even more.

The reader is advised on page 45 to put the larger mean square in the denominator of F. This is not consistent with the modern attitude toward the analysis of variance (e.g. Eisenhart, *Biometrics* 3 (1947), pp. 1–21).

On page 58 the reader has completed the first analysis of the tomato data. He has found that $\overline{A}$ "must be supposed to . . . (be) greater than 1," that $\overline{B}$ "does not . . . differ significantly from 1," and "that there is no reason to suppose that a real difference . . . exists (between $\overline{A}$ and $\overline{B}$)." (Parenthetic words ours.) The reviewer feels that the reader deserves to have the situation clarified.

The reader who reads on page 76 "is . . . appropriate . . . in this case, it being usual to take the highest order interaction for this purpose, unless . . . " will be surprised to learn nine pages later, on page 85, that he should use a different error term.

The reviewer cannot combine the comments on pages 78 and 131 on pooling estimates of error into a coherent philosophy. His own philosophy is not coherent either, but he would do the opposite in both cases. Perhaps Mather's third edition will clarify this problem for all of us.

The trout fly example on page 123 seems somewhat incautiously chosen—the distributions seem so far from normality as to make the author's concern about differences between significance at the 4%, 2% or 1% level somewhat exaggerated. The data thus do not make an important point with the definiteness the point deserves.

A major error of fact occurs on page 205, and again on page 212, where Mather states that maximum likelihood estimates have variances *equal* to $1/ni$ in samples of n. The example of the Cauchy distribution for $n=1$ or $n=2$ shows the falsity of this statement—here i is finite but the variance is infinite. It is known that as $n\to\infty$ the variance is *asymptotically* $1/ni$, but this is all.

On page 213 we are told that "It will be readily appreciated that inconsistent statistics are utterly misleading and should not be used under any circumstances." This seems exaggerated, in view of such inconsistent statistics for the mean of a normal distribution as the following: "Take the mean of the sample, or, if there are more observations than electrons in the metagalaxy, take twice the mean." (This is essentially due to Mosteller.)

On page 222 we find n turned into $n-1$ in a most *ad hoc* way. It is regrettably true that the maximum likelihood estimate of the variance of a normal population involves division by n, and not $n-1$. Regrettable, since this encourages n-dividers in their recreant ways. But since it is so, it should be admitted.

On page 239 we are told that the variance of $x^{1/2}$, x Poisson, is exactly independent of the mean. Asymptotic and exact statements are confounded here, as are the means of the population and sample.

JOHN W. TUKEY